IMM 361
September 1967

Courant Institute of
Mathematical Sciences

Contributions to Robust Estimation

Valerie Miké

Prepared under Contract Nonr-285(38) with
the Office of Naval Research NR 042-206

Distribution of this document is unlimited.

New York University

NEW YORK UNIVERSITY
COURANT INSTITUTE - LIBRARY
251 Mercer St. New York, N.Y. 10012

NR 042-206

IMM 361
September 1967

New York University

Courant Institute of Mathematical Sciences

CONTRIBUTIONS TO ROBUST ESTIMATION

Valerie Miké

This report represents results obtained at the Courant Institute of Mathematical Sciences, New York University, with the Office of Naval Research, Contract Nonr-285(38).
Reproduction in whole or in part is permitted for any purpose of the United States Government.

Distribution of this document is unlimited.

NEW YORK UNIVERSITY
COURANT INSTITUTE • LIBRARY

ABSTRACT

For the problem of efficiency-robust estimation of the location parameter θ of a family of symmetric pdf's $f(x-\theta|\lambda)$, $\lambda \in \Lambda = \{1,\ldots,m\}$, $\theta \in \Theta = \{\theta | -\infty < \theta < \infty\}$, the method of "mixture models" of Birnbaum (Ann. Math. Statist. 32 (1961) 622) is applied to determine generalized Pitman estimators, which are shown to be admissible, with squared error loss function, under broad regularity conditions. With increasing sample size, these estimators are proved to be fully efficient (i.e., asymptotically equivalent, for each value of λ, to the maximum likelihood estimator which would be appropriate if the true value of λ were known). Computationally tractable analogous estimators based on k sample quantiles are defined in the context of the model representing their asymptotic normal distributions. It is shown that with increasing k these approach equivalence to the fully efficient estimators based on complete samples. Equivalent estimators are given also for the case of unknown scale parameters. Efficiency-robust linear unbiased estimators based on sample quantiles are derived and the optimal spacing of quantiles is discussed.

ACKNOWLEDGEMENTS

The author is deeply grateful to Professor Allan Birnbaum who suggested this problem-area and provided constant guidance and encouragement during its investigation. Miss Bonnie Shinn contributed expert skills to typing the manuscript.

Page

1. INTRODUCTION 1

2. THE GENERALIZED PITMAN ESTIMATOR 9
   - 2.1. Introduction: The Pitman Estimator. . . . 9
   - 2.2. The Generalized Pitman Estimator. 13
   - 2.3. Asymptotic Properties 21

3. THE SYSTEMATIC GENERALIZED PITMAN ESTIMATOR. . . 36
   - 3.1. Introduction: Systematic Estimation . . . 36
   - 3.2. The Systematic Generalized Pitman Estimator: Exact Version. 40
   - 3.3. The Systematic Generalized Pitman Estimator: Approximate Version. . . . 44
   - 3.4. Asymptotic Properties of the Approximate Version 51
   - 3.5. Case When Scale Parameter is Unknown. . . 58
   - 3.6. Discussion. 60

4. SYSTEMATIC LINEAR ESTIMATORS COMPUTED FROM COVARIANCE MATRIX. 61
   - 4.1. Introduction. 61
   - 4.2. Systematic Linear Estimators Based on Exact Covariance Matrix 70
   - 4.3. Systematic Linear Estimators Based on Asymptotic Covariance Matrix. 74
   - 4.4. Discussion. 78

Page

5. SYSTEMATIC LINEAR ESTIMATORS BASED ON CONTINUOUS WEIGHT FUNCTIONS. 80
    5.1. Derivation of Jung's Result for a Single Shape Under Weakened (Necessary and Sufficient) Conditions. 80
    5.2. Direct Analogue for Systematic Estimation and Arbitrary Spacing. . . 89
    5.3. Application to Mixture Model. 97

6. THE PROBLEM OF OPTIMUM SPACING 103
    6.1. Ogawa's Optimum Solution for a Single Shape. 103
    6.2. Särndal's Nearly Optimum Solution for a Single Shape. 105
    6.3. Admissibility Concept for Mixture Model . 109
    6.4. Admissible Spacing Based on Exact Solution. 112
    6.5. Another Possible Approach 114
    6.6. Admissible Spacing Based on Approximate Solution. 116
    6.7. Discussion. 119

Chapter 1

INTRODUCTION

The term "robust" in the title of this paper refers to the broad program for development of efficiency-robust estimators initiated by Tukey (1960, 1962).

A general method for constructing optimally efficiency-robust inference techniques was given by Birnbaum (1961). In the chapters that follow we study the properties of various estimators of location obtained by this method.

We assume in this approach that we are given n independent, identically distributed random variables $X_1, \ldots, X_n$ with pdf $f(x - \theta|\lambda)$, where f is continuous and symmetric, $\theta \in \Theta = \{\theta | -\infty < \theta < \infty\}$, and $\lambda \in \Lambda = \{1, \ldots, m\}$ is a "shape" parameter; we are interested in finding estimators that are in some sense optimum uniformly over all shapes λ. Birnbaum's method yields estimators admissible over the parameter space $\Theta \times \Lambda$ with respect to some specified risk function. (A more detailed discussion of his method is given in Section 2.2 and also in Birnbaum and Laska (1967a).)

In Chapter 2 the efficiency-robust version of Pitman's (1939) estimator of location, the "generalized Pitman estimator," is examined. The admissibility of this estimator over $\Theta \times \Lambda$ was indicated by Birnbaum (1961). In Theorem 2.2 it is shown that, under mild regularity conditions, for fixed n the generalized Pitman estimator is unbiased, has variance independent of θ for every $\lambda \in \Lambda$, and is admissible over $\Theta \times \Lambda$ with variance as risk function.

The main result of the chapter is Theorem 2.6, in which we prove that under suitable regularity conditions the generalized Pitman estimator is asymptotically fully efficient under each shape λ; that is, it is asymptotically equivalent, under each shape, to the maximum likelihood estimator appropriate for that shape.

Since the Pitman estimator is in general quite difficult to compute even for a single shape, in Chapter 3 we consider an approximate version.

Mosteller (1946) proposed a method of estimation based on sample quantiles, in which, given a sample of size n, only a subset of size k is used in the construction of the estimators. After the observations have been ordered, the k sample quantiles $x_{(n_i)}$ ($i = 1, \ldots, k$; $x_{(n_i)} < x_{(n_{i+1})}$) to be selected are determined by the "spacing" $\{\lambda_i;\ i = 1, \ldots, k\}$, a set of k real numbers satisfying the relationship $0 < \lambda_1 < \ldots < \lambda_k < 1$, as follows: $n_i = n\lambda_i$ if $n\lambda_i$ is an integer, and $n_i = [n\lambda_i] + 1$ otherwise. We will refer to this method henceforth as that of systematic statistics.

A basic tool in related investigations is Mosteller's result that the k sample quantiles have an asymptotic k-variate normal distribution with mean and covariance matrix of known form.

What makes these estimators especially attractive is the fact that optimum properties are approached as both n and k tend to ∞, while k/n tends to 0. Extensive studies of some systematic estimators have been made by Ogawa (1951) and Särndal (1962) and in all cases investigated numerically, high efficiencies were obtained already for quite small values of k ($4 \leq k \leq 10$).

In Chapters 3 - 5 we consider the problem of constructing efficiency-robust systematic estimators. We assume that k and the spacing $\{\lambda_i\}$ are given, and that furthermore the spacing is symmetric, i.e., $\lambda_i = 1 - \lambda_{k-i+1}$, $i = 1, \ldots, k$.

In Chapter 3 we first define the systematic generalized Pitman estimator based on the exact distribution of the k selected order statistics. In the context of the latter model, we show that it has a number of optimum properties; but this estimator is still extremely difficult to compute.

We turn next to the systematic generalized Pitman estimator based on the asymptotic k-variate normal distribution of the sample quantiles; ease of computation is an immediate result. In Theorem 3.2 we prove that as the sample size n increases, this estimator is asymptotically efficient among all estimators based on the k-variate normal model, uniformly over all shapes λ. Following it we have in Theorem 3.3 the even stronger result that the asymptotic relative efficiency of this systematic generalized Pitman estimator in the class of all estimators of θ can be made arbitrarily close to unity (by increasing k) uniformly over all shapes λ. (This result refers to efficiency defined in the context of the exact model, for full samples.)

An asymptotically equivalent estimator is given in Section 3.5 for the case of practical importance in which the scale parameters σ_λ, $\lambda = 1, \ldots, m$, are unknown.

Chapter 4 deals with systematic linear estimators. Using the exact covariance matrix of the sample quantiles, we extend to the systematic case the best linear unbiased estimators (BLUE's) of Lloyd (1952) and, for efficiency-robustness, the admissible linear unbiased estimators (ALUE's) of Laska (1962), obtaining the corresponding best systematic linear unbiased estimators (BSLUE's) and admissible systematic linear unbiased estimators (ASLUE's) of θ.

If instead of the exact covariance matrix we consider the covariance matrix of the asymptotic normal distribution of the k sample quantiles, we have the asymptotically best systematic linear unbiased estimators (ABSLUE's) studied by Ogawa (1951), and their efficiency-robust counterparts, the asymptotically admissible systematic linear unbiased estimators (AASLUE's). The properties claimed for these estimators are proved, by methods analogous to the cases already given in the literature.

Next we show the rather useful result that the "asymptotic" estimators, the ABSLUE's and AASLUE's, are asymptotically equivalent to within $0_p(n^{-2})$ to the BSLUE's and ASLUE's, respectively. Finally, it follows from the work of Tischendorf (1955) that, as $k \longrightarrow \infty$, the ABSLUE's are also asymptotically efficient in the class of all estimators of θ.

In Chapter 5 we consider systematic linear estimators based on a continuous weight function. Jung (1955), among others, obtained an asymptotic version of Lloyd's best linear unbiased estimator which was generalized to the efficiency-robust case by Yhap (1966, 1967). Särndal (1962) gave the systematic version of Jung's result for an arbitrary spacing $\{\lambda_i\}$. In Section 5.1 we derive Jung's estimator under weakened (necessary and sufficient) conditions, and in Section 5.2 we show that Särndal's result, obtained by him by a different procedure, follows quite simply from Jung's approach.

If we consider the "uniform" spacing $\lambda_i = i/(k+1)$, $i = 1, \ldots, k$, the systematic weight function coincides with that of Jung. Hence, for this particular spacing, the efficiency-robust solution is the same for both the systematic and full-sample model. For this reason we give a brief account in Section 5.3 of Yhap's estimator. His approach seems to be adaptable to the case of an arbitrary spacing $\{\lambda_i\}$, and we indicate how such a solution would be attempted.

The reason that the uniform spacing is of real interest in Chapter 5 is that numerical studies have yielded good results for other than "optimum" spacings. The problem-area of the choice of k and the spacing $\{\lambda_i\}$ is discussed in Chapter 6. There we give a brief exposition of the method of optimum spacing defined by Ogawa (1951) and the "nearly optimum" solution proposed by Särndal (1962). We then define the concept of an admissible spacing for the efficiency-robustness problem and set up solutions analogous to those of Ogawa and Särndal. A third possible approach is also suggested.

All of the estimators here discussed are fully efficiency-robust in terms of the admissibility concept as defined in this context by Birnbaum (1961). In addition, the non-linear estimators treated in Chapters 2 and 3 have full asymptotic efficiency in the class of all estimators, uniformly over all shapes λ. Of course, since these results are asymptotic, numerical investigations are needed to determine how fast the various optimum properties approach their asymptotic values. In addition all these estimators, and others proposed for robust estimation, need to be compared numerically for small values of n and k, various spacings $\{\lambda_i\}$, and selected families of underlying distributions.

Huber (1964) defined a so-called (M)-estimator which he showed to be minimax with respect to asymptotic variance in the class of all translation-invariant estimators for a family of shapes C whose cdf's lie within a certain "distance" of the normal.

Concerning the relations between minimax estimators and possible fully efficient estimators, Huber has written (p. 80):

"At first glance, it might seem absurd to look for asymptotic minimax solutions, since, asymptotically, one could do better by estimating the true underlying distribution (see also Stein (1956), Hájek (1962)). However, this seems to require an exorbitant number of observations. On the other hand, it is hoped-- and preliminary numerical experiments seem to substantiate this hope--that (M)-estimators approach their asymptotic behavior rather fast, provided ψ is bounded. So an asymptotic minimax theory should be useful in those frequent cases where the sample size is perhaps large enough to indicate deviations from the assumed model but not yet large enough to establish their nature."

If we apply to the problem considered by Huber the systematic generalized Pitman estimator discussed in Chapter 3, we have to select a finite subset C' of symmetric shapes in C. (We consider only symmetric shapes; the usefulness and even the very meaning of the problem of estimating location of a family including various asymmetric shapes seems questionable, and in any case requires further clarification.) However, our estimator reaches full efficiency as k tends to ∞ for each member of C', whereas this is true only for a particular ("least favorable") shape in the case of the (M)-estimator. Since C' may be made arbitrarily large, it may be possible to choose it so as to obtain a very effective representation of C.

Furthermore, it seems not unreasonable to conjecture that the systematic generalized Pitman estimators, for small but carefully chosen families of shapes, may approach their asymptotic properties even for moderate sample sizes, since they are functions just of sample quantiles. The latter feature, incidentally, also makes these estimators highly tractable computationally, unlike some other estimators proposed for efficiency-robustness (including Huber's); this will facilitate the proposed numerical investigation of their properties, and their possible future applications.

Chapter 2

THE GENERALIZED PITMAN ESTIMATOR

2.1. Introduction: The Pitman Estimator

We shall assume henceforth that we have a fixed number n of independent, identically distributed random variables $X_1, \ldots, X_n$ with probability density $f(x|\theta) = f(x - \theta)$, $|\theta| < \infty$, where $f(x|\theta)$ is continuous in x and θ and is a symmetric function of x. Many of our results hold without the assumption of symmetry, but only the symmetric case is of ultimate relevance here. These assumptions on f will be referred to subsequently as Condition 1 or (C1).

(The underlying probability model is of course the usual one: A sample space $\mathcal{X}$, a σ-algebra $\mathcal{B}$ of subsets of $\mathcal{X}$, and a family of probability measures P_θ, $\theta \in \Omega$, defined over $(\mathcal{X}, \mathcal{B})$. In our case $\Omega = \Theta = R$, and $f(x|\theta)$ is the probability density of P_θ with respect to Lebesgue measure.)

In what follows we will use the simplified notation (x_i signifying an observation on X_i),

$$(x_1, \ldots, x_n) = \underline{x} \tag{2.1}$$

and

$$\prod_{i=1}^{n} f(x_i|\theta) = f(\underline{x}, \theta) .$$

Then if $xf(x) = O(1)$ as $x \longrightarrow \pm \infty$ and $x \longrightarrow 0$, a possible estimator of the location parameter θ is given by

$$(2.2) \qquad \theta^*(\underline{x}) = \frac{\int_{-\infty}^{\infty} \theta f(\underline{x},\theta)d\theta}{\int_{-\infty}^{\infty} f(\underline{x},\theta)d\theta} .$$

Pitman (1939), who proposed this estimator, proved that it (a) is translation-invariant, (b) has variance independent of θ, (c) is unbiased, and (d) has minimum variance among all translation-invariant estimators. Girshick and Savage (1951) showed in addition to (a) - (d) that, given certain regularity conditions, $\theta^*(\underline{x})$ is minimax with respect to squared error as loss in the class of all estimators. Karlin (1958) established its admissibility; an alternate proof under weaker conditions was given by Stein (1959) in the following theorem.

THEOREM 2.1. If

$$(2.3) \qquad \int_{R^n} f(\underline{x},\theta) \left\{ \frac{\int_{-\infty}^{\infty} \theta^2 f(\underline{x},\theta)d\theta}{\int_{-\infty}^{\infty} f(\underline{x},\theta)d\theta} - \left(\frac{\int_{-\infty}^{\infty} \theta f(\underline{x},\theta)d\theta}{\int_{-\infty}^{\infty} f(\underline{x},\theta)d\theta} \right)^2 \right\}^{3/2} \prod_{i=1}^{n} dx_i < \infty ,$$

then $\theta^*(\underline{x})$ *as defined by* (2.2) *is admissible. That is, if* $\hat{\theta}(\underline{x})$ *is any estimator for which*

$$E_\theta[\hat{\theta}(\underline{x}) - \theta]^2 \leq E_\theta[\theta^*(\underline{x}) - \theta]^2 \quad \text{for all} \quad \theta ,$$

then equality holds for all θ.

For future reference we indicate here the general setting in which the proof is carried out.

Let $\mathcal{Y}$ be the (n-1)-dimensional real coordinate space, $\mathcal{C}$ the σ-algebra of Borel subsets of $\mathcal{Y}$, and ν the distribution of Y defined by

$$(2.4) \qquad g(y) = \frac{\int xf(x)f(x+y_1) \ \ldots \ f(x+y_{n-1})dx}{\int f(x)f(x+y_1) \ \ldots \ f(x+y_{n-1})dx} \ ,$$

where $y = (y_1, \ldots, y_{n-1})$, and X is a random variable on $(R,\mathcal{B})$. Define the mapping

$$(2.5) \qquad \begin{aligned} Y_1 &= X_2 - X_1 \\ &\vdots \\ Y_{n-1} &= X_n - X_1 \\ X &= X_1 - g(Y_1, \ldots, Y_{n-1}) \ . \end{aligned}$$

Then the conditional density of X given Y is given by

$$(2.6) \qquad p(x,y) = \frac{f(x+g(y))f(x+g(y)+y_1) \ \dots \ f(x+g(y)+y_{n-1})}{\int f(x)f(x+y_1) \ \dots \ f(x+y_{n-1})dx} ,$$

X is the estimator $\theta^*(\underline{x})$, and the condition for admissibility (2.2) reduces to

$$(2.7) \qquad \int d\nu(y)\left(\int x^2 p(x,y)dx\right)^{3/2} < \infty .$$

In subsequent sections we will refer to (2.3), Stein's condition for the admissibility of the Pitman estimator (which implies its existence), simply as (C2).

2.2 The Generalized Pitman Estimator

Birnbaum (1961) has given a general method for constructing efficiency-robust inference techniques. Application of this method to the Pitman estimator is discussed in Birnbaum and Laska (1967a).

A brief summary of the method as it pertains to our problem follows.

Assume given a family of m densities $f(x - \theta,\lambda)$, $\lambda \in \Lambda = \{1,\ldots, m\}$, satisfying (C1) and (C2). The additional parameter here signifies that the observations $x_1,\ldots, x_n$ may have come from any one of the m distributions or "shapes" defined by $f(x - \theta,\lambda)$. θ is the unknown parameter of interest, and λ plays the role of a nuisance parameter; thus the parameter space for this model is $\Omega = \Theta \times \Lambda$.

For the efficiency-robustness problem, realistic assumptions include continuous families of shapes; the present model is a simplified one whose ultimate value and possible limitations require further investigation. (Cf. also discussion in Birnbaum and Laska (1967a, b)).

Let $T = T(\underline{x})$ be an estimator of θ and let $R = R(T,\theta,\lambda)$ denote any specified risk function, e.g. mean squared error. Admissibility of T with respect to R in a specified class of estimators is defined in the usual way; namely, if there exists an estimator $T' = T'(\underline{x})$ in the given class such that

$$R(T',\theta,\lambda) \leq R(T,\theta,\lambda), \quad \forall\, (\theta,\lambda) \in \Omega ,$$

then equality holds $\forall (\theta,\lambda) \in \Omega$.

Since we assume that a smaller value of R indicates greater efficiency, the object is to find a T that somehow "jointly" minimizes $R(T,\theta,\lambda)$, $(\theta,\lambda) \in \Omega$. In general no T is uniformly best over Ω.

Given a probability distribution G over Λ, i.e., $G = (g_1,\ldots, g_m)$, $\sum_{\lambda=1}^{m} g_\lambda = 1$, $g_\lambda > 0$, we introduce the so-called "mixture model"

$$\begin{aligned} f(x,\theta,G) &= \int_\Lambda f(x,\theta,\lambda)dG(\lambda) \\ &= \sum_{\lambda=1}^{m} g_\lambda f(x,\theta,\lambda) \ . \end{aligned} \tag{2.8}$$

The risk function of an estimator T under the mixture model is then given by

$$\begin{aligned} R(T(\underline{x}),\theta,G) &= \int_{R^n} L(T,\theta) \prod_{i=1}^{n} f(x_i,\theta,G)dx_i \\ &= \sum_{\lambda=1}^{m} g_\lambda \int_{R^n} L(T,\theta) \prod_{i=1}^{n} f(x_i,\theta,\lambda)dx_i \\ &= \sum_{\lambda=1}^{m} g_\lambda \ R(T(\underline{x}),\theta,\lambda) \ , \end{aligned} \tag{2.9}$$

where $L(T,\theta)$ is the loss function corresponding to R. We now have the following

LEMMA 2.1. *Any estimator* $T = T(\underline{x})$ *admissible over* Θ *under the mixture model is also admissible in the usual sense over* $\Theta \times \Lambda$.

Proof: Assume $T(\underline{x})$ is inadmissible over $\Theta \times \Lambda$ with respect to R in a given class of estimators. Then $\exists\ T' = T'(\underline{x})$, $\ni$

$$R(T',\theta,\lambda) \leq R(T,\theta,\lambda), \quad \forall (\theta,\lambda) \in \Theta \times \Lambda$$

and $\exists\ (\theta',\lambda') \in \Theta \times \Lambda \ni$

$$R(T',\theta',\lambda') < R(T,\theta',\lambda') .$$

Now $\forall\ \theta \neq \theta'$

$$\begin{aligned} R(T',\theta,G) &= \sum_{\lambda=1}^{m} g_\lambda R(T',\theta,\lambda) \\ &\leq \sum_{\lambda=1}^{m} g_\lambda R(T,\theta,\lambda) \\ &= R(T,\theta,G) \end{aligned}$$

and

$$\begin{aligned} R(T',\theta',G) &= \sum_{\lambda \neq \lambda'} g_\lambda R(T',\theta',\lambda) + g_{\lambda'} R(T',\theta',\lambda') \\ &< \sum_{\lambda=1}^{m} g_\lambda R(T,\theta',\lambda) \\ &= R(T,\theta',G) . \end{aligned}$$

But this contradicts the assumption that T is admissible under (2.8).

It needs to be emphasized that the "prior" distribution G employed in the mixture model has no extra-mathematical significance and is selected in any particular context because it yields an admissible estimator with additional desirable properties.

Clearly estimators admissible in the sense defined above also fit the description "fully efficiency-robust," as discussed in Chapter 1.

Now, for any given G, computing the Pitman estimator for the mixture model (2.8) yields the generalized Pitman estimator

$$\theta^*_G(\underline{x}) = \frac{\int_{-\infty}^{\infty} \theta f(\underline{x},\theta,G)\,d\theta}{\int_{-\infty}^{\infty} f(\underline{x},\theta,G)\,d\theta} \tag{2.10}$$

$$= \frac{\sum_{\lambda=1}^{m} g_\lambda \int_{-\infty}^{\infty} \theta f(\underline{x},\theta,\lambda)\,d\theta}{\sum_{\lambda=1}^{m} g_\lambda \int_{-\infty}^{\infty} f(\underline{x},\theta,\lambda)\,d\theta}$$

$$= \sum_{\lambda=1}^{m} \frac{g_\lambda \int_{-\infty}^{\infty} f(\underline{x},\theta,\lambda)\,d\theta}{\sum_{j=1}^{m} g_j \int_{-\infty}^{\infty} f(\underline{x},\theta,j)\,d\theta} \cdot \theta^*_\lambda(\underline{x})$$

$$= \sum_{\lambda=1}^{m} g^*_\lambda \theta^*_\lambda ,$$

where $\theta^*_\lambda = \theta^*_\lambda(\underline{x})$ is the Pitman estimator for shape λ, and $g^*_\lambda = g^*_\lambda(\underline{x})$ is a formal "posterior probability" of shape λ.

The following results are immediate.

THEOREM 2.2. The generalized Pitman estimator θ^*_G defined by (2.10)

(a) is unbiased

(b) has variance independent of $\theta \; \forall \; \lambda$, and

(c) is admissible (over $\Theta \times \Lambda$).

Proof: (a) and (b) follow from the fact that the regularity conditions assumed for each shape $f(x - \theta, \lambda)$ carry over to the mixture model $f(x - \theta, G)$, so that the results established by Pitman (1939) and Girshick and Savage (1951) for a single shape are immediately applicable.

(c) is a special case of Lemma 2.1. We have

$$\text{(2.11)} \qquad \operatorname{Var}(\theta^*_G | G) = \int_{R^n} (\theta^* - \theta)^2 \prod_{i=1}^{n} f(x_i, G) dx_i$$

$$= \sum_{\lambda=1}^{m} g_\lambda \int_{R^n} (\theta^*_G - \theta)^2 \prod_{i=1}^{n} f(x_i, \lambda) dx_i$$

$$= \sum_{\lambda=1}^{m} g_\lambda \operatorname{Var}(\theta^*_G | \lambda) \; .$$

Now assume θ^*_G is inadmissible. Then there exists $\hat{\theta} = \hat{\theta}(\underline{x})$ such that

$$\text{(2.12)} \qquad \operatorname{Var}(\hat{\theta} | \theta, \lambda) \leq \operatorname{Var}(\theta^*_G | \lambda) \; , \qquad \forall \; (\lambda, \theta)$$

with inequality strict for at least one (λ,θ), say (λ',θ'). Then, for $\theta \neq \theta'$,

$$(2.13) \qquad \mathrm{Var}(\hat{\theta}|\theta,G) = \sum_{\lambda\neq\lambda'} g_\lambda \, \mathrm{Var}(\hat{\theta}|\theta,\lambda) + g_{\lambda'} \, \mathrm{Var}(\hat{\theta}|\theta,\lambda')$$

$$\leq \sum g_\lambda \, \mathrm{Var}(\theta^*_G|\lambda)$$

$$= \mathrm{Var}(\theta^*_G|G) \ ,$$

and

$$(2.14) \qquad \mathrm{Var}(\hat{\theta}|\theta',G) = \sum_{\lambda\neq\lambda'} g_\lambda \, \mathrm{Var}(\hat{\theta}|\theta',\lambda) + g_{\lambda'} \, \mathrm{Var}(\hat{\theta}|\theta',\lambda)$$

$$< \sum g_\lambda \, \mathrm{Var}(\theta^*_G|\lambda)$$

$$= \mathrm{Var}(\theta^*_G|G) \ ,$$

which contradicts the admissibility of θ^*_G over $\mathcal{H}$ (under the mixture model) following from Theorem 2.1. This proves the theorem.

As an example of the generalized Pitman estimator, consider the case of $m = 2$, that is, 2 possible shapes, with $\lambda = 1, 2$ and $G = (g,1-g)$.

We have

$$(2.15) \qquad \theta^*_G = \frac{g \int \theta f(\underline{x},\theta,1)d\theta + (1-g) \int \theta f(\underline{x},\theta,2)d\theta}{g \int f(\underline{x},\theta,1)d\theta + (1-g) \int f(\underline{x},\theta,2)d\theta}$$

$$= \frac{gA + (1-g)B}{gC + (1-g)D} \ .$$

Let

$$f(x - \theta,1) = \frac{1}{\sqrt{2\pi}} \exp\{- \tfrac{1}{2} (x - \theta)^2 \}$$

and

$$f(x - \theta,2) = \frac{1}{\sqrt{2}} \exp\{- \sqrt{2}\, |x - \theta| \},$$

the normal and double exponential densities with unit variance.

We obtain, for sample size n,

$$A = \frac{\bar{x}}{(2\pi)^{(n-1)/2}\sqrt{n}} \exp[- \tfrac{1}{2} \sum_{i=1}^{n} (x_i - \bar{x})^2]$$

and

$$C = \frac{1}{(2\pi)^{(n-1)/2}\sqrt{n}} \exp[- \tfrac{1}{2} \sum_{i=1}^{n} (x_i - \bar{x})^2] .$$

In the following expressions for B and D the undefined terms $\sum_{i=1}^{0} x_i$ and $\sum_{i=n+1}^{n} x_i$ are to be set equal to zero, $x_o = - \infty$ and $x_{n+1} = + \infty$. We have, for n odd,

$$B = \frac{1}{2^{(n+1)}\sqrt{2}} \sum_{j=0}^{n} \frac{\exp[\sqrt{2} (\sum_{i=1}^{j} x_i - \sum_{i=j+1}^{n} x_i)]}{2j - n}$$

$$\cdot \Big\{ [x_j + \frac{1}{\sqrt{2}(2j-n)}] \exp[- \sqrt{2} (2j - n)x_j]$$

$$- [x_{j+1} + \frac{1}{\sqrt{2}(2j-n)}] \exp[- \sqrt{2} (2j - n)x_{j+1}] \Big\} ,$$

and for n even,

$$B = \frac{1}{2^{(n+1)/2}} \sum_{\substack{j=0 \\ j \neq \frac{n}{2}}}^{n} \frac{\exp[\sqrt{2}\,(\sum_{i=1}^{j} x_i - \sum_{i=j+1}^{n} x_i)]}{2j - n}$$

$$\cdot \left\{ [x_j + \frac{1}{\sqrt{2}(2j-n)}]\exp[-\sqrt{2}(2j-n)x_j] \right.$$

$$\left. - [x_{j+1} + \frac{1}{\sqrt{2}(2j-n)}]\exp[-\sqrt{2}(2j-n)x_{j+1}] \right\}$$

$$+ \frac{1}{2^{n/2+1}}[x^2_{\frac{n}{2}+1} - x^2_{\frac{n}{2}}]\exp[\sqrt{2} \sum_{i=1}^{n/2} x_i - \sum_{i=\frac{n}{2}+1}^{n} x_i] .$$

For n odd,

$$D = \frac{1}{2^{(n+1)/2}} \sum_{j=0}^{n} \frac{\exp[\sqrt{2}\,(\sum_{i=1}^{j} x_i - \sum_{i=j+1}^{n} x_i)]}{2j - n}$$

$$\cdot \{ \exp[-\sqrt{2}(2j - n)x_j] - \exp[-\sqrt{2}(2j - n)x_{j+1}] \} ,$$

and for n even,

$$D = \frac{1}{2^{(n+1)/2}} \sum_{\substack{j=0 \\ j \neq \frac{n}{2}}}^{n} \frac{\exp[\sqrt{2}\,(\sum_{i=1}^{j} x_i - \sum_{i=j+1}^{n} x_i)]}{2j - n}$$

$$\cdot \{ \exp[-\sqrt{2}(2j - n)x_j] - \exp[-\sqrt{2}(2j - n)x_{j+1}] \}$$

$$+ \frac{1}{2^{n/2}}[x_{\frac{n}{2}+1} - x_{\frac{n}{2}}]\exp[\sqrt{2}\,(\sum_{i=1}^{n/2} x_i - \sum_{i=\frac{n}{2}+1}^{n} x_i)] .$$

2.3. Asymptotic Properties

If as "prior density" of θ we admit any non-negative measurable function, the Pitman estimator may be considered as the mean of the posterior distribution of θ given X for the case of the "generalized uniform distribution" (i.e., Lebesgue measure) over the real line. In this context it is also the so-called generalized Bayes solution, with squared error as loss.

The asymptotic normality of the posterior distribution for arbitrary smooth priors and binomial X was proved by Bernstein (1917) and independently by von Mises (1919). For the special case of a uniform prior, it had already been proved by Laplace (1820), under rather strong assumptions. That the Bernstein-von Mises result holds for general smooth $f(x|\theta)$ was shown by Wolfowitz (1953). These and related results are discussed in the rigorous study of the asymptotic properties of Bayes solutions and maximum likelihood estimators carried out by LeCam (1953, 1958). Additional work includes publications by Lindley (1961, 1965), Hartigan (1965), and Elfving (1966).

From the extensive literature on the subject we will make use only of the single result that, under broad regularity conditions, the mean of the posterior distribution--in our case the Pitman estimator for a single shape--is asymptotically equivalent to the corresponding maximum likelihood estimator.

Specifically, we have

$$\theta^*(\underline{x}) = \hat{\theta}(\underline{x}) + O_p(\tfrac{1}{n}) \;. \tag{2.16}$$

DEFINITION. Given a sequence of random variables $\{\xi_n\}$, we write

$$\xi_n = O_p(r_n) \tag{2.17}$$

i.e., $|\xi_n|/r_n$ is bounded in probability, if
$\forall \varepsilon > 0, \exists\, M, N \ni \forall\, n > N$

$$P\{|\xi_n| \geq Mr_n\} < \varepsilon \;.$$

We are adopting here the general notation of Chernoff (1956).

A sketch of (2.16) will be given below, in the form needed subsequently. The method involves expansion of $\log f(x,\theta)$ in Taylor series about $\hat{\theta}(\underline{x})$. Following Lindley (1961) we will consider only $(\theta - \hat{\theta}) = O_p(n^{-1/2})$ and will integrate over the entire range of θ. It can be shown, using, e.g., methods of Hartigan (1965) and Elfving (1966), that the omitted integrals over the tail areas in question are negligible.

PRELIMINARIES. The following conditions are assumed to hold for the underlying density $f(x - \theta)$; let $\theta = \theta_0 \in \Theta$ be the true value of the parameter. (These conditions are not all necessary in this form; a more precise formulation will be worked out later.)

(C2). The Pitman estimator exists and is admissible, $\forall$ n.

(C3). $(\partial^r/\partial\theta^r) \log f(x - \theta)$, $1 \leq r \leq 4$, exists $\forall$ x and θ.

(C4). $|E^{(r)}(\theta,\theta')| = |E_\theta[(\partial^r/\partial\theta^r) \log f(X - \theta)|_{\theta = \theta'}]| < \infty$,

$$1 \leq r \leq 4, \quad \forall\, \theta,\ \theta' .$$

(C5). $|E(\theta,\theta')| = |E_\theta \log f(X - \theta')| < \infty, \forall\, \theta,\ \theta'$.

(C6). $0 < I(\theta,\theta') < \infty, \forall\, \theta,\ \theta'$,

where

$$I(\theta,\theta') = E_\theta[(\partial/\partial\theta) \log f(X - \theta)|_{\theta = \theta'}]^2 ,$$

and

$$I(\theta,\theta') = - E^{(2)}(\theta,\theta') .$$

(Here $nI(\theta_o,\theta_o)$ is Fisher's Information.)

(C7). The maximum likelihood estimator $\hat{\theta}(\underline{x})$ is the unique solution of the likelihood equation for $n \geq N$, for some N and $\forall$x; and the MLE is asymptotically $N(\theta_o, 1/nI(\theta_o,\theta_o))$.

When considering the family of densities $f(x,\theta,\lambda)$, $\lambda \in \Lambda$, we assume that in addition to the above, the following conditions hold. Let $\lambda = \lambda_o \in \Lambda$ be the true value of the shape parameter.

(C5.a). $|E(\theta,\theta';\lambda,\lambda')| = |E_{(\theta,\lambda)}[\log f(X - \theta'|\lambda')]| < \infty$,

$$\forall\, (\theta,\lambda),\ (\theta',\lambda') \in \Omega .$$

(C6.a). $0 < I(\theta,\theta';\lambda,\lambda') < \infty$, $\forall (\theta,\lambda), (\theta',\lambda') \in \Omega$,

where

$$I(\theta,\theta';\lambda,\lambda') = E_{(\theta,\lambda)}[(\partial/\partial\theta) \log f(X - \theta|\lambda')|_{\theta = \theta'}]^2 ,$$

and

$$I(\theta,\theta';\lambda,\lambda') = - E^{(2)}(\theta,\theta';\lambda,\lambda') .$$

(We note that $nI(\theta_o,\theta_o;\lambda_o,\lambda_o)$ is Fisher's Information for shape λ_o.)

(C7.a). $\hat{\theta}_\lambda(\underline{x})$, the unique root of the likelihood equation determined by shape λ, is asymptotically $N(\theta_o, 1/nI(\theta_o,\theta_o;\lambda_o,\lambda_o))$.

When referring to the expressions defined above we will adopt the simpler notation of writing when appropriate e.g.

$$E(\theta,\theta) = E(\theta)$$
$$E^{(r)}(\theta,\theta) = E^{(r)}(\theta)$$
$$I(\theta,\theta) = I(\theta)$$

$$I(\theta,\theta';\lambda,\lambda) = I(\theta,\theta';\lambda)$$
$$I(\theta,\theta;\lambda,\lambda') = I(\theta;\lambda,\lambda')$$
$$I(\theta,\theta;\lambda,\lambda) = I(\theta;\lambda) ,$$

and similarly for $E(\theta,\theta';\lambda,\lambda')$.

(C7) and (C7.a) are satisfied by the normal, logistic, and double exponential distributions, but not, e.g., by the Cauchy. Under weakened assumptions admitting the Cauchy distribution, one could allow multiple roots of the likelihood equation such that the probability of more than one root $\longrightarrow 0$ with n; this case will not be discussed here.

We note, in addition, that since $f(x - \theta)$ belongs to the translation-parameter family, the functions in (C4)-(C7) are actually independent of θ for arguments of the form (θ,θ), and depend only on $(\theta - \theta')$ when the argument is (θ,θ').

Repeated use will be made of two theorems which we state here without proof.

THEOREM 2.3. (Wilks (1963), p. 105). *Let* $(X_1, X_2,\dots)$ *be a stochastic process depending on a parameter* θ *such that for each value of* θ *in some interval* (θ',θ''), $f_n(x_1,\dots, x_n, \theta)$, $n = 1, 2,\dots$, *and* $\theta_n^*(x_1,\dots, x_n)$, $n = 1, 2,\dots$, *are sequences of random variables converging in probability respectively to* $g(\theta)$ *and* θ *uniformly with respect to* θ *in* (θ',θ''), *where* $g(\theta)$ *is continuous in* (θ',θ''). *Then* $f_n(x_1,\dots, x_n, \theta_n^*)$, $n = 1, 2,\dots$, *converges in probability to* $g(\theta)$.

THEOREM 2.4. (Slutsky; e.g. p. 102 in Rao (1965)). *Let* (X_n,Y_n), $n = 1, 2,\dots$ *be a sequence of pairs of random variables. Then*

$$\text{(a)}\quad X_n \xrightarrow{\mathcal{L}} X,\ Y_n \xrightarrow{P} 0 \quad\Longrightarrow\quad X_nY_n \xrightarrow{P} 0$$

$$\text{(b)}\quad X_n \xrightarrow{\mathcal{L}} X,\ Y_n \xrightarrow{P} c \quad\Longrightarrow\quad X_nY_n \xrightarrow{\mathcal{L}} cX$$

$$\Longrightarrow\quad X_n + Y_n \xrightarrow{\mathcal{L}} X + c\ .$$

In addition we will need the following

LEMMA 2.2. Assume $f(x - \theta)$ satisfies (C2)-(C7) and let

$$(2.18) \qquad L_r(\underline{x}|\theta) = \sum_{i=1}^{n} \frac{\partial^r}{\partial\theta^r} \log f(x_i - \theta), \quad 1 \leq r \leq 4.$$

Then

$$(2.19) \qquad L_r(\underline{x}|\hat{\theta}(\underline{x})) = O_p(n),$$

where $\hat{\theta}(\underline{x})$ is the maximum likelihood estimator.

Proof: By (C4) $E^{(r)}(\theta_o,\theta)$ exists, so that, by the weak law of large numbers,

$$(2.20) \qquad \frac{1}{n} L_r(\underline{x}|\theta) \xrightarrow{P} E^{(r)}(\theta_o,\theta) < \infty .$$

By the consistency of the maximum likelihood estimator, we have

$$(2.21) \qquad \hat{\theta}(\underline{x}) \xrightarrow{P} \theta_o .$$

Application of Theorem 2.3 now yields

$$(2.22) \qquad \frac{1}{n} L_r(\underline{x}|\hat{\theta}) \xrightarrow{P} E^{(r)}(\theta_o) < \infty ,$$

i.e.,

$$L_r(\underline{x}|\hat{\theta}) = O_p(n) .$$

THEOREM 2.5. Assume given a sequence of independent random variables $X_1, X_2, \ldots,$ with p.d.f. $f(x - \theta)$ satisfying

(C1)-(C7). Then the Pitman estimator may be written

$$(2.23) \qquad \theta^*(\underline{x}) = \frac{\int_{-\infty}^{\infty} \theta f(\underline{x},\theta)d\theta}{\int_{-\infty}^{\infty} f(\underline{x},\theta)d\theta}$$

$$= \hat{\theta}(\underline{x}) + o_p(\tfrac{1}{n}) .$$

Proof: We consider first the denominator; for simplicity, we write

$$(2.24) \qquad L(\underline{x}|\theta) = \log f(\underline{x}|\theta) = \sum_{i=1}^{n} \log f(x_i|\theta)$$

and

$$L_r(\underline{x}|\theta) = \frac{\partial^r}{\partial\theta^r} \log f(\underline{x}|\theta) ,$$

then expand $L(\underline{x}|\theta)$ in a Taylor series about $\hat{\theta}$. We obtain

$$(2.25) \qquad L(\underline{x}|\theta) = L(\underline{x}|\hat{\theta}) + \frac{1}{2} (\theta - \hat{\theta})^2 L_2(\underline{x}|\hat{\theta})$$

$$+ \frac{1}{3!} (\theta - \hat{\theta})^3 L_3(\underline{x}|\hat{\theta}) + \frac{1}{4!} (\theta - \hat{\theta})^4 L_4(\underline{x}|\theta')$$

for some θ' in the interval $(\hat{\theta} - \theta, \hat{\theta} + \theta)$. By (2.20)

$$\frac{1}{n} L_4(\underline{x}|\theta') \xrightarrow{P} E^{(4)}(\theta_o,\theta') < \infty$$

or

$$L_4(\underline{x}|\theta') = O_p(n) ,$$

so that we have

$$(2.26)\quad f(\underline{x}|\theta) = f(\underline{x}|\hat{\theta}) \exp\{\tfrac{1}{2}(\theta - \hat{\theta})^2 L_2(\underline{x}|\hat{\theta}) + \tfrac{1}{3!}(\theta - \hat{\theta})^3 L_3(\underline{x}|\hat{\theta}) + 0_p(n^{-1})\}$$

$$= f(\underline{x}|\hat{\theta}) \exp\{\tfrac{1}{2}(\theta - \hat{\theta})^2 L_2(\underline{x}|\hat{\theta})\}[1 + + \tfrac{1}{3!}(\theta - \hat{\theta})^3 L_3(\underline{x}|\hat{\theta}) + 0_p(n^{-1}) + \tfrac{1}{2}[t(\theta)]^2 \exp\{\delta t(\theta)\}] .$$

where

$$0 < \delta < 1$$

and

$$(2.27)\qquad t(\theta) = \tfrac{1}{3!}(\theta - \hat{\theta})^3 L_3(\underline{x}|\hat{\theta}) + 0_p(n^{-1}) .$$

(Let $t(\theta) = y$; we have substituted above $e^y = 1 + y + \frac{1}{2} y^2 e^{\delta y}$, valid for arbitrary y.) Calculating the order of the remainder, we obtain

$$(2.28)\quad \tfrac{1}{2}[t(\theta)]^2 \exp\{\delta t(\theta)\} = 0_p(n^{-1})0(\exp\{n^{-1/2}\}) = 0_p(n^{-1}) ,$$

since by assumption $(\theta - \hat{\theta}) = 0_p(n^{-1/2})$. Thus we have

$$(2.29)\quad \int f(\underline{x}|\theta)d\theta = f(\underline{x}|\hat{\theta}) \int d\theta \exp\{\tfrac{1}{2}(\theta - \hat{\theta})^2 L_2(\underline{x}|\hat{\theta})\} \cdot [1 + \tfrac{1}{3!}(\theta - \hat{\theta})^3 L_3(\underline{x}|\hat{\theta}) + 0_p(n^{-1})] .$$

Integration with respect to θ yields

$$(2.30) \quad \int f(\underline{x}|\theta)d\theta = f(\underline{x}|\hat{\theta})\left[\frac{-2\pi}{L_2(\underline{x}|\hat{\theta})}\right]^{1/2}[1 + O_p(\tfrac{1}{n})] \; ,$$

where by (C6), requiring a unique root of the likelihood equation, $L_2(\underline{x}|\hat{\theta}) < 0$.

Next, consider the numerator of (2.23). Write

$$(2.31) \quad \int \theta f(\underline{x}|\theta)d\theta = \hat{\theta}\int f(\underline{x}|\theta)d\theta + \int (\theta - \hat{\theta})f(\underline{x}|\theta)d\theta \; .$$

Since we have already evaluated the first integral on the right-hand side, only the last term needs to be examined. Completely analogously to the previous arguments we obtain

$$(2.32) \quad \int f(\underline{x}|\theta)d\theta = f(\underline{x}|\hat{\theta})\int d\theta \, \exp\{\tfrac{1}{2}(\theta - \hat{\theta})^2 L_2(\underline{x}|\hat{\theta})\}$$

$$\cdot \, [(\theta - \hat{\theta}) + \tfrac{1}{3!}(\theta - \hat{\theta})^4 L_3(\underline{x}|\hat{\theta}) + O_p(n^{-3/2})]$$

$$= f(\underline{x}|\hat{\theta})\left[\frac{-2\pi}{L_2(\underline{x}|\hat{\theta})}\right]^{1/2}\left[\frac{3L_3(\underline{x}|\hat{\theta})}{2\cdot 3!L_2^2(\underline{x}|\hat{\theta})} + O_p(n^{-3/2})\right]$$

$$= f(\underline{x}|\hat{\theta})\left[\frac{-2\pi}{L_2(\underline{x}|\hat{\theta})}\right]^{1/2}\left[O_p(n^{-1})\right] \; .$$

Substituting (2.30), (2.31), and (2.32) in (2.23) yields

$$(2.33) \qquad \theta^* = \hat{\theta} + \frac{O_p(n^{-1})}{1+O_p(n^{-1})} = \hat{\theta} + O_p(n^{-1}) \; .$$

COROLLARY 2.1. *Under the assumptions of Theorem* 2.5 *the Pitman estimator has asymptotic distribution* $N(\theta_o, 1/nI(\theta_o))$, *where* θ_o *is the true value of* θ *and* $nI(\theta_o)$ *is Fisher's Information. That is,* θ^* *is a BAN estimator.*

Proof: Since by assumption $\hat{\theta}$ is a BAN estimator, the result follows from part (b) of Theorem 2.4, with $c = 0$.

We now come to the main result of this section, namely, that as $n \longrightarrow \infty$ the generalized Pitman estimator is not only admissible, but has full asymptotic efficiency.

THEOREM 2.6. *Let* $f(x,\theta,\lambda)$, $\lambda = 1, 2, \ldots, m$, *be* m *densities satisfying* (C1)-(C7), *and let* $\theta = \theta_o$ *and* $\lambda = \lambda_o$ *be the true parameter values. If* $\forall\ (\theta,\lambda) \neq (\theta_o,\lambda_o)$ $f(x,\theta,\lambda)$ *differs from* $f(x,\theta_o,\lambda_o)$ *on more than a set of measure zero, then the generalized Pitman estimator* $\theta^*_G(\underline{x})$ *has asymptotic distribution* $N(\theta_o, 1/nI(\theta_o;\lambda_o))$, *where* $I(\theta_o;\lambda_o)$ *is Fisher's Information for shape* λ_o, *that is,* θ^*_G *is BAN uniformly over the respective shapes.*

COMMENT. Familiar examples show that the maximum likelihood estimators for individual shapes are in general BAN only under a single shape. For example, we have the normal and double exponential shapes, for which the maximum likelihood estimators are the mean and the median, respectively, neither of which is uniformly BAN over both shapes.

Proof: The form of the generalized Pitman estimator was given by (2.10). Substituting for the integrals therein the asymptotic expressions (2.30) - (2.31), with the appropriate addition of the indices λ, we obtain

$$(2.34) \qquad \theta_G^* = \sum_{\lambda=1}^{m} \frac{g_\lambda \int_{-\infty}^{\infty} f(\underline{x}|\theta,\lambda)d\theta}{\sum_{j=1}^{m} g_j \int_{-\infty}^{\infty} f(\underline{x}|\theta,j)d\theta} \; \theta_\lambda^*$$

$$= \sum_{\lambda=1}^{m} \frac{g_\lambda f(\underline{x}|\hat{\theta}_\lambda,\lambda)\left[\dfrac{-2\pi}{L_2(\hat{\theta}_\lambda|\lambda)}\right]^{1/2}}{\sum_{j=1}^{m} g_j f(\underline{x}|\hat{\theta}_j,j)\left[\dfrac{-2\pi}{L_2(\hat{\theta}_j|j)}\right]^{1/2}} \left[\hat{\theta}_\lambda + O_p(n^{-1})\right],$$

where

$$L_2(\hat{\theta}_\lambda|\lambda) = L_2(\underline{x}|\hat{\theta}_\lambda,\lambda) \; .$$

Without loss of generality assume that $\lambda_o = 1$ is the true value of the shape parameter. Then (2.34) can be written as

$$(2.35) \qquad \theta_G^* = \frac{1}{1 + \sum_{j=2}^{m} \dfrac{g_j f(\underline{x}|\hat{\theta}_j,j)}{g_1 f(\underline{x}|\hat{\theta}_1,1)}\left[\dfrac{L_2(\hat{\theta}_1|1)}{L_2(\hat{\theta}_j|j)}\right]^{1/2}} \left[\hat{\theta}_1 + O_p(n^{-1})\right]$$

$$+ \sum_{\lambda=2}^{m} \frac{\dfrac{g_\lambda f(\underline{x}|\hat{\theta}_\lambda,\lambda)}{g_1 f(\underline{x}|\hat{\theta}_1,1)}\left[\dfrac{L_2(\hat{\theta}_1|1)}{L_2(\hat{\theta}_\lambda|\lambda)}\right]^{1/2}}{1 + \sum_{j=2}^{m} \dfrac{g_j f(\underline{x}|\hat{\theta}_j,j)}{g_1 f(\underline{x}|\hat{\theta}_1,1)}\left[\dfrac{L_2(\hat{\theta}_1|1)}{L_2(\hat{\theta}_j|j)}\right]^{1/2}} \left[\hat{\theta}_\lambda + O_p(n^{-1})\right].$$

Now by the consistency of the maximum likelihood estimator, (cf. (C7.a)),

$$\hat{\theta}_\lambda \equiv \hat{\theta}_\lambda(\underline{x}) \xrightarrow{P} \theta_o \ , \ \forall \ \lambda \in \Lambda \ . \tag{2.36}$$

By the weak law of large numbers and by (C6.a)

$$\begin{aligned} \frac{1}{n} L_2(\underline{x}|\theta,\lambda) &\xrightarrow{P} E^{(2)}(\theta_o,\theta;1,\lambda) \\ &= E_{(\theta_o,1)}\left[\frac{\partial^2}{\partial\theta^2} \log f(X,\theta,\lambda)\right] , \\ &\qquad \forall \lambda \in \Lambda \ . \end{aligned} \tag{2.37}$$

Application of Theorem 2.3 now yields

$$\frac{1}{n} L_2(\hat{\theta}_\lambda|\lambda) \xrightarrow{P} E^{(2)}(\theta_o;1,\lambda), \ \forall \ \lambda \in \Lambda \ ,$$

so that

$$\left[\frac{L_2(\hat{\theta}_1|1)}{L_2(\hat{\theta}_\lambda|\lambda)}\right]^{1/2} \xrightarrow{P} \left[\frac{I(\theta_o;1)}{I(\theta_o;1,\lambda)}\right]^{1/2} = c_\lambda \ , \qquad 0 < c_\lambda < \infty \ . \tag{2.38}$$

Next we will show that

$$\frac{f(\underline{x}|\hat{\theta}_\lambda,\lambda)}{f(\underline{x}|\hat{\theta}_1,1)} \xrightarrow{P} 0, \quad \forall \ \lambda \neq 1 \ . \tag{2.39}$$

Let

(2.40) $$u_n = \frac{f(\underline{x}|\hat{\theta}_\lambda,\lambda)}{f(\underline{x}|\hat{\theta}_1,1)} = \prod_{i=1}^{n} \frac{f(x_i|\hat{\theta}_\lambda,\lambda)}{f(x_i|\hat{\theta}_1,1)} ,$$

and let

(2.41) $$v_n = \frac{1}{n} \log u_n$$

$$= \frac{1}{n}\sum_{i=1}^{n} \log f(x_i|\hat{\theta}_\lambda,\lambda) - \frac{1}{n}\sum_{i=1}^{n} \log f(x_i|\hat{\theta}_1,1) .$$

By the weak law of large numbers and by (C5.a)

(2.42) $$\frac{1}{n}\sum_{i=1}^{n} \log f(x_i,\theta_o,\lambda) \xrightarrow{P} E(\theta_o;1,\lambda)$$

$$= E_{(\theta_o,1)}[\log f(X|\theta_o,\lambda)], \quad \forall \lambda \in \Lambda .$$

Now from (2.36) and Theorem 2.3 we obtain

(2.43) $$\frac{1}{n}\sum_{i=1}^{n} \log f(x_i|\hat{\theta}_\lambda,\lambda) \xrightarrow{P} E_{(\theta_o,1)}[\log f(X|\theta_o,\lambda)] ,$$

$$\forall \lambda \in \Lambda ,$$

so that

(2.44) $$v_n \xrightarrow{P} E_{(\theta_o,1)}\left[\log \frac{f(X,\theta_o,\lambda)}{f(X,\theta_o,1)}\right] = E_1(v) .$$

Now, following Hartigan (1965) we have, by the convexity of the exponential function,

(2.45) $$\exp[E_1(v)] \leq E_1[\exp v] = 1,$$

since

$$E_1[\exp v] = \int_{-\infty}^{\infty} \frac{f(x,\theta_o,\lambda)}{f(x,\theta_o,1)} f(x,\theta_o,1)dx = 1.$$

The equality holds only if $v \equiv c$ with probability 1. But by assumption $f(x,\theta_o,\lambda)$ differs from $f(x,\theta_o,1)$ with probability > 0, so that v is not constant and

$$\exp[E_1(v)] < 1 \tag{2.46}$$

or

$$E_1(v) = E_{(\theta_o,1)}\left[\log \frac{f(X,\theta_o,\lambda)}{f(X,\theta_o,1)}\right] < 0 \ ;$$

that is,

$$v_n \xrightarrow{P} E_1(v) = -k, \quad k > 0,$$

so that

$$nv_n \xrightarrow{P} -\infty$$

and

$$u_n = e^{nv_n} \xrightarrow{P} 0 \ . \tag{2.47}$$

Repeated application of Theorem 2.4 now completes the proof of the theorem: Let

$$\theta_G^* = d_1[\hat{\theta}_1 + 0_p(n^{-1})] + \sum_{\lambda=2}^{m} d_\lambda[\hat{\theta}_\lambda + 0_p(n^{-1})] \tag{2.48}$$

$$= d_1\hat{\theta}_1 + \sum_{\lambda=2}^{m} d_\lambda\hat{\theta}_\lambda + 0_p(n^{-1}) \ ,$$

where d_λ, $\lambda \in \Lambda$, are the coefficients shown in (2.35) and

$$\sum_{\lambda=1}^{m} d_\lambda = 1.$$

(d_λ is the approximate form of the coefficient g_λ^* defined in (2.10).) From (2.38) and (2.39) we see that

(2.49) $$d_1 \equiv d_1(\underline{x}) \xrightarrow{P} 1$$

and

$$d_\lambda \equiv d_\lambda(\underline{x}) \xrightarrow{P} 0 , \qquad \lambda = 2, \ldots, m.$$

By assumption (C7.a)

$$\hat{\theta}_\lambda \xrightarrow{\mathcal{L}} Z_\lambda \sim N(\theta_o, \frac{1}{nI(\theta_o;1,\lambda)}) .$$

Thus

$$d_1\hat{\theta}_1 \xrightarrow{\mathcal{L}} Z_1 \sim N(\theta_o, \frac{1}{nI(\theta_o;1)}) ,$$

$$d_\lambda\hat{\theta}_\lambda \xrightarrow{P} 0 ,$$

and

$$O_p(n^{-1}) \xrightarrow{P} 0 ,$$

so that

(2.50) $$\theta_G^* \xrightarrow{\mathcal{L}} Z_1 \sim N(\theta_o, 1/nI(\theta_o;1)) ,$$

as claimed.

Chapter 3

THE SYSTEMATIC GENERALIZED PITMAN ESTIMATOR

3.1. Introduction: Systematic Estimation

Mosteller (1946) proposed a method of estimation for large samples based on only a subset of the observations. It involves replacing the observations $x_1, \ldots, x_n$ by the ordered set of their values,

$$x_{(1)} < x_{(2)} < \ldots < x_{(n)} ,$$

and then for a fixed set of k integers n_i, $i = 1, \ldots, k$, selecting the set of sample quantiles $x_{(n_1)}, \ldots, x_{(n_k)}$, such that

$$x_{(n_1)} < x_{(n_2)} < \ldots < x_{(n_k)} .$$

The sample quantiles are determined by the relation

$$(3.1) \qquad n_i = \begin{cases} n\lambda_i & \text{if } n\lambda_i \text{ is an integer} \\ [n\lambda_i]+1 & \text{otherwise} \end{cases}$$

where $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_k < 1$ is a given set of k real numbers.

Although Mosteller defines any function of the ordered observations as a "systematic" statistic, we will restrict the latter term to the method actually introduced by him, i.e., the use of only a selected set of sample quantiles.

Mosteller's proposal was motivated by the coming into use of punched-card equipment which could provide an economical way of sorting (ordering) large masses of data. Tedious further computations would then be restricted to the small selected subset.

A significant contribution in this area was his proof that under mild regularity conditions the k sample quantiles $X_{(n_1)}, \ldots, X_{(n_k)}$ have as limiting distribution the k-variate normal, with mean and covariance matrix of known form. (This result had been proved earlier for $k = 1, 2$ by K. Pearson and Smirnoff; for references see Mosteller (1946).)

As a first example Mosteller suggested using the mean of the selected subsample as an estimator of location and he computed the relative efficiency of this estimator based on different sets $\{\lambda_i, i = 1, \ldots, k\}$, $k \leq 10$, when the underlying distribution was assumed normal. In each case the variance of the estimator was computed from the covariance matrix of the asymptotic k-variate normal distribution and compared with the variance of the full-sample mean. For all spacings $\{\lambda_i\}$ considered, the relative efficiency exceeded .9 already for $k = 5$.

Ogawa (1951), in an extensive study of this approach, applied the generalized Gauss-Markov theorem to the covariance matrix of the limiting multivariate normal distribution, obtaining so-called best linear unbiased estimators (BLUE's) for this k-variate model, pre-dating the well-known paper of Lloyd (1952). He also studied the problem of optimum spacing of the sample quantiles and showed (as was to be expected) that the BLUE's based on these quantiles provided relative efficiencies for the normal distribution even more favorable than those obtained by Mosteller.

Extensions of Ogawa's work were given by Tischendorf (1955). Ogawa also contributed several expository chapters to Sarhan and Greenberg (1962). A comprehensive summary of the whole problem area, with additional new results, is given in the treatise by Särndal (1962).

These references, all pertaining to linear estimators (not taking into consideration here applications to the theory of testing), will be discussed further in subsequent chapters.

Although data processing has come a long way since the early marvel of sorting machines, the method of using sample quantiles is by no means obsolete. More sophisticated methods of constructing estimators can give rise to elaborate problems of numerical analysis even when k, the number of sample quantiles used, is quite small relative to the sample size n.

In addition, an important application has been found in the space program. Given a large set of univariate measurements made by a satellite, it is sometimes economical to have only a subset sent back to earth, unless all computations are performed aboard the spacecraft. It has been shown that it is cheaper to include the necessary equipment for ordering the observations and selecting the subsample than to provide a more elaborate computing system. For a full discussion of this problem, see, e.g., Eisenberger and Posner (1965), and references cited therein.

In view of the above considerations, and the fact that low values of k already yielded very high relative efficiencies in the cases studied, the problem of "systematic" estimation seems to warrant further investigation, particularly in connection with problems of efficiency-robust estimation, to which we turn next.

3.2. The Systematic Generalized Pitman Estimator: Exact Version

Given n identically distributed random variables $X_1, \ldots, X_n$ with pdf $f(x - \theta, \lambda)$ satisfying conditions (C1)-(C7) for $\lambda = 1, \ldots, m$, and k real numbers λ_i, $0 < \lambda_1 < \lambda_2 < \ldots < \lambda_k < 1$, let $X_{(n_1)}, \ldots, X_{(n_k)}$ be a subset of the variables, with n_i as defined in (3.1). Assume that the spacings are symmetric, i.e.,

$$\lambda_i = 1 - \lambda_{k-i+1} , \quad i = 1, \ldots, k . \tag{3.2}$$

In what follows, let

$$\underline{x}^{(k)} = (x_{(n_1)}, \ldots, x_{(n_k)})$$

and denote by $w(\underline{x}^{(k)} | \theta, \lambda)$ the joint pdf of the k sample quantiles under shape λ.

Then the generalized Pitman estimator based on this model may be written

$$\theta_G^*(k,n) = \frac{\sum_{\lambda=1}^{m} g_\lambda \int_{-\infty}^{\infty} \theta w(\underline{x}^{(k)} | \theta, \lambda) d\theta}{\sum_{\lambda=1}^{m} g_\lambda \int_{-\infty}^{\infty} w(\underline{x}^{(k)} | \theta, \lambda) d\theta} \tag{3.3}$$

and, from (4.2) of Section 4.1,

$$\theta_G^*(k,n) = \frac{\sum_{\lambda=1}^{m} g_\lambda \int_{-\infty}^{\infty} \theta \prod_{i=1}^{k+1} [F(x_{(n_i)}-\theta|\lambda)-F(x_{(n_{i-1})}-\theta|\lambda)]^{n_i-n_{i-1}-1}}{\sum_{\lambda=1}^{m} g_\lambda \int_{-\infty}^{\infty} \prod_{i=1}^{k+1} [F(x_{(n_i)}-\theta|\lambda)-F(x_{(n_{i-1})}-\theta|\lambda)]^{n_i-n_{i-1}-1}}$$

$$\cdot \frac{\prod_{i=1}^{k} f(x_{(n_i)}-\theta|\lambda)d\theta}{\prod_{i=1}^{k} f(x_{(n_i)}-\theta|\lambda)d\theta} ,$$

where

$$n_o = 0 , \quad n_{k+1} = n+1 ,$$

$$x_{(n_o)} = -\infty , \quad x_{(n_{k+1})} = +\infty ,$$

and

$$F(x|\lambda) = \int_{-\infty}^{x} f(t|\lambda)dt .$$

All properties established in Theorem 2.2 for the generalized Pitman estimator apply within the context of this model:

(a) Unbiasedness follows from the additional assumption that the spacings are symmetric.

(b) The variance of $\theta_G^*(k,n)$ is independent of θ because this estimator, too, is translation invariant, i.e.,

$$\theta_G^*(\underline{x}^{(k)} + \underline{c}|k,n) = \theta_G^*(\underline{x}^{(k)}|k,n) + c , \tag{3.4}$$

where

$$\underline{x}^{(k)} + \underline{c} = (x_{(n_1)} + c, \ldots, x_{(n_k)} + c)$$

and c is an arbitrary constant. We have, from (3.3)

$$\theta^*_G(\underline{x}^{(k)}+\underline{c}\,|\,k,n)$$

$$= \frac{\sum_{\lambda=1}^{m} g_\lambda \int_{-\infty}^{\infty} \theta \prod_{i=1}^{k+1} \left[F(x_{(n_i)}-(\theta-c)|\lambda)-F(x_{(n_{i-1})}-(\theta-c)|\lambda)\right]^{n_i-n_{i-1}-1}}{\sum_{\lambda=1}^{m} g_\lambda \int_{-\infty}^{\infty} \prod_{i=1}^{k+1} \left[F(x_{(n_i)}-(\theta-c)|\lambda)-F(x_{(n_{i-1})}-(\theta-c)|\lambda)\right]^{n_i-n_{i-1}-1}}$$

$$\cdot \frac{\prod_{i=1}^{k} f(x_{(n_i)}-(\theta-c)|\lambda)\,d\theta}{\prod_{i=1}^{k} f(x_{(n_i)}-(\theta-c)|\lambda)\,d\theta} \quad .$$

Performing the change of variable

$$\theta' = \theta - c$$

immediately yields (3.4).

The method of proof of part (c) of Theorem 2.2 is also applicable, since Stein's method of proving the admissibility of the Pitman estimator, Theorem 2.1, can similarly be applied in this context: Let

$$x_{(n_i)} = x_i \ , \qquad i = 1,\ldots, k \ ,$$

and

$$(3.5) \qquad g(y) = \frac{\int_{-\infty}^{\infty} x w(x,x+y_1,\ldots, x+y_{k-1})dx}{\int_{-\infty}^{\infty} w(x,x+y_1,\ldots, x+y_{k-1})dx} \quad .$$

Then the transformation (2.5) yields, as the conditional density of X given Y,

$$(3.6) \qquad p(x,y) = \frac{w(x+f(y),\ x+f(y)+y_1, \ldots,\ x+f(y)+y_{k-1})}{\int_{-\infty}^{\infty} w(x,\ x+y_1, \ldots,\ x+y_{k-1})\,dx} ,$$

and the proof of Theorem 2.1 proceeds as before.

3.3. The Systematic Generalized Pitman Estimator: Approximate Version

Although the estimator $\theta_G^*(k,n)$, based on k sample quantiles, inherits all the desirable properties of the full-sample Pitman estimator within the context of its own model, it is evidently also extremely difficult to compute. However, the fact that the sample quantiles have a limiting multivariate normal distribution suggests the possibility of defining a corresponding estimator based on this asymptotic model. One immediate result is ease of computation.

This new estimator, denoted $\theta_G^*(k)$, will be introduced in this section, and its properties corresponding to Theorem 2.2 established.

But how good is an estimator based on an approximate model? We will find, in Section 3.4, that not only is $\theta_G^*(k)$ BAN, as $n \longrightarrow \infty$, in the class of estimators based on the k-variate normal model, but also that with increasing k it achieves full asymptotic efficiency in the class of all estimators, uniformly over all shapes λ.

We state here, for ready reference, Mosteller's theorem concerning the asymptotic distribution of the k order statistics.

THEOREM 3.1. (e.g. Sarhan and Greenberg (1962), p. 17). Let $X_1, \ldots, X_n$ be n independent, identically distributed random variables with pdf $f(x|\theta)$. For k given real numbers for which

$$0 = \lambda_0 < \lambda_1 < \lambda_2 < \ldots < \lambda_k < \lambda_{k+1} = 1$$

let the λ_i*-quantile of the distribution be* x_i, *that is*

$$(3.7) \qquad \int_{-\infty}^{x_i} f(t|\theta)dt = \lambda_i , \quad i = 1,\ldots, k .$$

Assume that $f(x)$ *is differentiable in the neighborhood of* $x = x_i$, $i = 1,\ldots, k$, *and let*

$$f_i = f(x_i) \neq 0 \quad \text{for} \quad i = 1,\ldots, k .$$

Then the joint distribution of the k *sample quantiles*

$$X_{(n_1)}, X_{(n_2)}, \ldots, X_{(n_k)} ,$$

where

$$n_i = \begin{cases} n\lambda_i & \text{if } n\lambda_i \text{ is an integer} \\ [n\lambda_i]+1 & \text{otherwise} \end{cases}$$

tends to a k-*dimensional normal distribution with means* $x_1,\ldots, x_k$ *and with covariance matrix*

$$\frac{1}{n} \Gamma = \frac{1}{n} [\frac{\lambda_i(1-\lambda_j)}{f_i f_j}] , \quad i \leq j , \quad \Gamma \text{ symmetric}$$

as $n \longrightarrow \infty$.

When f depends only on a location parameter θ, i.e., is of the form $f(x - \theta)$, let the λ_i-quantile of the standardized distribution be u_i, that is,

$$(3.8) \qquad \int_{-\infty}^{u_i} f(y)dy = \lambda_i \ ,$$

so that

$$(3.9) \qquad E[X_{(n_i)}] = x_i = u_i + \theta \ .$$

(For a derivation of (3.9), and other related results, see Section 4.1.) Then the pdf of the limiting distribution of the k sample quantiles $X_{(n_1)}, \ldots, X_{(n_k)}$ is

$$(3.10) \quad h(\underline{x}^{(k)}|\theta)$$

$$= (\tfrac{n}{2\pi})^{k/2} \, |\Gamma|^{-1/2} \exp\{-\tfrac{n}{2}(\underline{x}^{(k)}-\underline{\theta}-\underline{u})'\Gamma^{-1}(\underline{x}^{(k)}-\underline{\theta}-\underline{u})\} \ ,$$

where

$\underline{\theta} = (\theta, \ldots, \theta)$, a k-vector,
$\underline{u} = (u_1, \ldots, u_k)$,
$|\Gamma| = \det \Gamma$,

$$\Gamma = \begin{bmatrix} \frac{\lambda_1(1-\lambda_1)}{f_1^2} & \frac{\lambda_1(1-\lambda_2)}{f_1 f_2} & \cdots & \frac{\lambda_1(1-\lambda_k)}{f_1 f_k} \\ \frac{\lambda_1(1-\lambda_2)}{f_1 f_2} & \frac{\lambda_2(1-\lambda_2)}{f_2^2} & \cdots & \frac{\lambda_2(1-\lambda_k)}{f_2 f_k} \\ \vdots & & & \\ \frac{\lambda_1(1-\lambda_k)}{f_1 f_k} & \frac{\lambda_2(1-\lambda_k)}{f_2 f_k} & \cdots & \frac{\lambda_k(1-\lambda_k)}{f_k^2} \end{bmatrix} ;$$

and $f_i = f(u_i)$, $i = 1, \ldots, k$; that is,

$$(3.11) \quad h(\underline{x}^{(k)} | \theta)$$

$$= \left(\frac{n}{2\pi}\right)^{k/2} f_1 \ldots f_k [\lambda_1(\lambda_2-\lambda_1)\ldots(\lambda_k-\lambda_{k-1})(1-\lambda_k)]^{-1/2}$$

$$\cdot \exp\left\{- \frac{n}{2} \left[\sum_{i=1}^{k} \frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})} f_i^2 (x_{(n_i)}-\theta-u_i)^2\right.\right.$$

$$\left.\left. - 2 \sum_{i=2}^{k} \frac{f_i f_{i-1}}{\lambda_i-\lambda_{i-1}} (x_{(n_i)}-\theta-u_i)(x_{(n_{i-1})}-\theta-u_{i-1})\right]\right\} .$$

In what follows we assume that the spacings satisfy (3.2). Then for $k < n$ and Γ non-singular, we define the approximate systematic Pitman estimator

$$(3.12) \quad \theta^*(k) = \frac{\int_{-\infty}^{\infty} \theta h(\underline{x}^{(k)} | \theta) d\theta}{\int_{-\infty}^{\infty} h(\underline{x}^{(k)} | \theta) d\theta}$$

$$= \frac{\int_{-\infty}^{\infty} \theta \exp\{- \frac{n}{2} (\underline{x}^{(k)}-\underline{u}-\underline{\theta})' S (\underline{x}^{(k)}-\underline{u}-\underline{\theta}) d\theta}{\int_{-\infty}^{\infty} \exp\{- \frac{n}{2} (\underline{x}^{(k)}-\underline{u}-\underline{\theta})' S (\underline{x}^{(k)}-\underline{u}-\underline{\theta}) d\theta} ,$$

where $S = [s_{ij}] = \Gamma^{-1}$, all other terms being defined in (3.10). For simplicity we will write $x_{(n_i)} = x_i$. Writing out the denominator, we have

$$(3.13)\quad \int_{-\infty}^{\infty} \exp\{-\tfrac{n}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}(x_i-u_i-\theta)(x_j-u_j-\theta)s_{ij}\}d\theta$$

$$= \int_{-\infty}^{\infty} \exp\{-\tfrac{n}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}[(x_i-u_i)(x_j-u_j)s_{ij}$$

$$-\theta(x_i-u_i+x_j-u_j)s_{ij} + \theta^2 s_{ij}]\}d\theta .$$

Completing the square, we obtain for (3.13)

$$(3.14)\quad \exp\left\{-\frac{n}{2}\left[\sum_{i=1}^{k}\sum_{j=1}^{k}(x_i-u_i)(x_j-u_j)s_{ij} - \frac{\left[\sum_{i=1}^{k}\sum_{j=1}^{k}(x_i-u_i+x_j-u_j)s_{ij}\right]^2}{4\sum_{i=1}^{k}\sum_{j=1}^{k}s_{ij}}\right]\right\}$$

$$\cdot \int_{-\infty}^{\infty} \exp\left\{-\frac{n}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}s_{ij}\left[\theta - \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}(x_i-u_i+x_j-u_j)s_{ij}}{2\sum_{i=1}^{k}\sum_{j=1}^{k}s_{ij}}\right]^2\right\}$$

$$= \left[\begin{array}{c}\text{expression up to integral}\\ \text{in (3.14)}\end{array}\right] \cdot \sqrt{\frac{2\pi}{n\sum_{i=1}^{k}\sum_{j=1}^{k}s_{ij}}} .$$

Completely analogously we obtain for the numerator of (3.12)

$$(3.15)\quad \int_{-\infty}^{\infty} \theta \exp\{-\frac{n}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}(x_i-u_i-\theta)(x_j-u_j-\theta)s_{ij}\}d\theta$$

$$= \left[\text{as above}\right]\cdot\sqrt{\frac{2\pi}{n\sum_{i=1}^{k}\sum_{j=1}^{k}s_{ij}}}\cdot\frac{\sum_{i=1}^{k}\sum_{j=1}^{k}(x_i-u_i+x_j-u_j)s_{ij}}{2\sum_{i=1}^{k}\sum_{j=1}^{k}s_{ij}},$$

so that (3.12) reduces to

$$(3.16)\qquad \theta^*(k) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k}(x_i-u_i+x_j-u_j)s_{ij}}{2\sum_{i=1}^{k}\sum_{j=1}^{k}s_{ij}}.$$

It is easy to verify that also

$$(3.17)\qquad \theta^*(k) = \hat{\theta}(k),$$

the maximum likelihood estimator, and

$$(3.18)\qquad \theta^*(k) = \tilde{\theta}(k) = \frac{1'S\underline{x}^{(k)'}}{1'S1},$$

the least squares estimator of θ corresponding to the given k-variate normal distribution. (Here $1 = (1,\ldots,1)'$, a k-vector, and we assume that (C1) holds.)

Let us assume now that we are given a family of m shapes $f(x-\theta,\lambda)$, $\lambda = 1,\ldots,m$, all satisfying (C1) and the conditions of Theorem 3.1. Supplying the index λ where necessary, we have immediately the corresponding systematic generalized Pitman estimator $\theta_G^*(k)$; that is,

$$(3.19)\quad \theta_G^*(k) = \sum_{\lambda=1}^{m} \frac{g_\lambda\left(\frac{|S_\lambda|}{1'S_\lambda 1}\right)^{1/2}\exp\{-\frac{n}{2}\,(\underline{x}^{(k)}-\underline{u}^\lambda-\underline{\hat{\theta}}_\lambda)'S_\lambda(\underline{x}^{(k)}-\underline{u}^\lambda-\underline{\hat{\theta}}_\lambda)\}}{\sum_{j=1}^{m} g_j\left(\frac{|S_j|}{1'S_j 1}\right)^{1/2}\exp\{-\frac{n}{2}\,(\underline{x}^{(k)}-\underline{u}^j-\underline{\hat{\theta}}_j)'S_j(\underline{x}^{(k)}-\underline{u}^j-\underline{\hat{\theta}}_j)\}} \cdot \hat{\theta}_\lambda(k)$$

where

$$S_\lambda = \Gamma_\lambda^{-1} ,$$

with Γ_λ the covariance matrix of the k-variate normal distribution for shape λ,

$$\hat{\theta}_\lambda = \text{MLE for shape } \lambda ,$$

$$\underline{\hat{\theta}}_\lambda = (\hat{\theta}_\lambda, \ldots, \hat{\theta}_\lambda), \text{ a k-vector},$$

$$\underline{u}^\lambda = E_\lambda[\underline{X}^{(k)}] - \underline{\theta} ,$$

$$1'S_\lambda 1 = \sum_{i=1}^{k} \sum_{j=1}^{k} s_{ij}^\lambda ,$$

and, as always, $\sum g_\lambda = 1,\ 0 < g_\lambda < 1, \forall\ \lambda$.

All comments in Section 3.2 concerning properties of $\theta_G^*(k,n)$ apply to $\theta_G^*(k)$ within the context of this model. In particular, $\theta_G^*(k)$ is admissible (with respect to $\Omega = \Theta \times \Lambda$, in the class of all estimators of θ) on the hypothetical assumption that the respective asymptotic distributions of the k selected sample quantiles $h(\underline{x}^{(k)}|\theta,\lambda),\ \lambda \in \Lambda$, represent the true model. As mentioned earlier, the sense in which this hypothetical assumption is also a pertinent approximation will be examined in the next section.

3.4. Asymptotic Properties of the Approximate Version

As before, we will assume without loss of generality that $\lambda = 1$ is the true value of the shape parameter. Then (3.19) may be rewritten as

$$(3.20)\quad \theta_G^*(k) = \frac{1}{1 + \sum_{j=2}^{m} \frac{g_j}{g_1}\left(\frac{|S_j|}{|S_1|} \cdot \frac{(1'S_1 1)}{(1'S_j 1)}\right)^{1/2}} \cdot \frac{1}{\frac{\exp\{-\frac{n}{2}(\underline{x}^{(k)}-\underline{u}^{j}-\hat{\underline{\theta}}_j)'S_j(\underline{x}^{(k)}-\underline{u}^{j}-\hat{\underline{\theta}}_j)\}}{\exp\{-\frac{n}{2}(\underline{x}^{(k)}-\underline{u}^{1}-\hat{\underline{\theta}}_1)'S_1(\underline{x}^{(k)}-\underline{u}^{1}-\hat{\underline{\theta}}_1)\}}} \cdot \hat{\theta}_1(k)$$

$$+ \sum_{\lambda=2}^{m} \frac{\frac{g_\lambda}{g_1}\left(\frac{|S_\lambda|}{|S_1|} \cdot \frac{(1'S_1 1)}{(1'S_\lambda 1)}\right)^{1/2}}{1 + \sum_{j=2}^{m} \frac{g_j}{g_1}\left(\frac{|S_j|}{|S_1|} \cdot \frac{(1'S_1 1)}{(1'S_j 1)}\right)^{1/2}}$$

$$\cdot \frac{\frac{\exp\{-\frac{n}{2}(\underline{x}^{(k)}-\underline{u}^{\lambda}-\hat{\underline{\theta}}_\lambda)'S_\lambda(\underline{x}^{(k)}-\underline{u}^{\lambda}-\hat{\underline{\theta}}_\lambda)\}}{\exp\{-\frac{n}{2}(\underline{x}^{(k)}-\underline{u}^{1}-\hat{\underline{\theta}}_1)'S_1(\underline{x}^{(k)}-\underline{u}^{1}-\hat{\underline{\theta}}_1)\}}}{\frac{\exp\{-\frac{n}{2}(\underline{x}^{(k)}-\underline{u}^{j}-\hat{\underline{\theta}}_j)'S_j(\underline{x}^{(k)}-\underline{u}^{j}-\hat{\underline{\theta}}_j)\}}{\exp\{-\frac{n}{2}(\underline{x}^{(k)}-\underline{u}^{1}-\hat{\underline{\theta}}_1)'S_1(\underline{x}^{(k)}-\underline{u}^{1}-\hat{\underline{\theta}}_1)\}}} \cdot \hat{\theta}_\lambda(k) .$$

We will first consider the limiting distribution of $\theta_G^*(k)$ as $n \longrightarrow \infty$ but k remains fixed.

We begin with

LEMMA 3.1. For $\lambda \neq 1$, $\underline{u}^{\lambda} \neq \underline{u}^{1}$, $S_{\lambda} \neq S_1$,

$$(3.21) \quad u_n(\lambda) = \frac{\exp\{-\frac{n}{2}(\underline{x}^{(k)}-\underline{u}^{\lambda}-\hat{\underline{\theta}}_{\lambda})'S_{\lambda}(\underline{x}^{(k)}-\underline{u}^{\lambda}-\hat{\underline{\theta}}_{\lambda}\}}{\exp\{-\frac{n}{2}(\underline{x}^{(k)}-\underline{u}^{1}-\hat{\underline{\theta}}_{1})'S_{1}(\underline{x}^{(k)}-\underline{u}^{1}-\hat{\underline{\theta}}_{1})\}} \xrightarrow{P} 0 .$$

Proof: We have $\underline{X}^{(k)} = (X_{(n_1)}, \ldots, X_{(n_k)})$, where

$$(3.22) \qquad E_{\lambda}[X_{(n_i)}] = u_i^{\lambda} + \theta .$$

Now

$$(3.23) \qquad \mathrm{Var}[X_{(n_i)}] = \frac{\lambda_i(1-\lambda_i)}{n[f(u_i^{\lambda}|\lambda)]^2} \longrightarrow 0$$

$$\Longrightarrow X_{(n_i)} \xrightarrow{P} E_{\lambda}[X_{(n_i)}] , \quad \lambda \in \Lambda ,$$

so that, in our case,

$$X_{(n_i)} \xrightarrow{P} u_i^1 + \theta .$$

From (3.18) we have

$$\hat{\theta}_{\lambda} = \frac{1'S_{\lambda}\underline{x}^{(k)'}}{1'S_{\lambda}1} ,$$

so that

$$(3.24) \qquad \mathrm{Var}[\hat{\theta}_{\lambda}|1] = \frac{1'S_{\lambda}\Gamma_1'S_{\lambda}1}{n(1'S_{\lambda}1)^2} \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty .$$

Thus $\hat{\theta}_{\lambda}$ too is consistent and goes in probability to θ with increasing n, since

(3.25) $$E_1\left(\frac{1'S_\lambda \underline{x}^{(k)'}}{1'S_\lambda 1}\right) = \frac{1'S_\lambda \underline{u}^1}{1'S_\lambda 1} + \theta = \theta \ .$$

That $1'S_\lambda \underline{u}^1 = 0$ was shown by Lloyd (cf. p. 24 of Sarhan and Greenberg (1962)). Thus we have

(3.26) $$\underline{x}^{(k)} - \underline{u}^1 - \hat{\underline{\theta}}_1 \xrightarrow{P} \underline{0}$$
and
$$\underline{x}^{(k)} - \underline{u}^\lambda - \hat{\underline{\theta}}_\lambda \xrightarrow{P} \underline{u}^1 - \underline{u}^\lambda \ .$$

Write (3.21) as

(3.27) $$u_n(\lambda) = \exp\{-\tfrac{n}{2}[(\underline{x}^{(k)}-\underline{u}^\lambda-\hat{\underline{\theta}}_\lambda)'S_\lambda(\underline{x}^{(k)}-\underline{u}^\lambda-\hat{\underline{\theta}}_\lambda) - (\underline{x}^{(k)}-\underline{u}^1-\hat{\underline{\theta}}_1)'S_1(\underline{x}^{(k)}-\underline{u}^1-\hat{\underline{\theta}}_1)]\}$$
$$= \exp\{-\tfrac{n}{2} v_n(\lambda)\} \ .$$

But by (3.26)

(3.28) $$v_n(\lambda) \xrightarrow{P} (\underline{u}^1 - \underline{u}^\lambda)'S_\lambda(\underline{u}^1 - \underline{u}^\lambda) > 0 \ ,$$

so that we have immediately

(3.29) $$u_n(\lambda) \xrightarrow{P} 0 \ .$$

This brings us to

THEOREM 3.2. $\theta^*_G(k)$ _is distributed asymptotically as_ $N(\theta, \frac{1}{n1'S_1 1})$, _where, as before,_ $\lambda = 1$ _is the true value of the shape parameter._

Proof: Write (3.20) as

$$\theta_G^*(k) = d_1\hat{\theta}_1(k) + \sum_{\lambda=2}^{m} d_\lambda\hat{\theta}_\lambda(k) \tag{3.30}$$

or

$$\sqrt{n}(\theta_G^*(k) - \theta) = d_1\sqrt{n}(\hat{\theta}_1(k) - \theta) + \sum_{\lambda=2}^{m} d_\lambda\sqrt{n}(\hat{\theta}_\lambda - \theta) \ . \tag{3.31}$$

From the lemma we have

$$\begin{aligned} & d_1 \equiv d_1(\underline{x}^{(k)}) \xrightarrow{P} 1 \\ \text{and} \quad & \\ & d_\lambda \equiv d_\lambda(\underline{x}^{(k)}) \xrightarrow{P} 0 \ , \quad \lambda = 2,\ldots, m. \end{aligned} \tag{3.32}$$

But, $\forall\ \lambda$,

$$\sqrt{n}(\hat{\theta}_\lambda - \theta) \sim N\left(0, \frac{1'S_\lambda\Gamma_1 S_\lambda 1}{n(1'S_\lambda 1)^2}\right), \quad \Gamma_1 = S_1^{-1},$$

so that application of Theorem 2.4 yields

$$\sqrt{n}(\theta_G^*(k) - \theta) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{n1'S_1 1}\right) \ ,$$

as claimed.

These results show that as the sample size n increases, $\theta_G^*(k)$ is asymptotically efficient among all estimators based on the k-variate normal model, uniformly over all shapes λ.

But, letting k increase also, we have an even stronger result.

THEOREM 3.3. *Assume that* $f(x - \theta|\lambda)$ *is everywhere differentiable with respect to* $u = x - \theta$, *for every* $\lambda \in \Lambda$, *and that* $0 < I(\theta;\lambda) < \infty$. *Then* $\theta^*_G(k)$ *is an* ε*-efficient estimator uniformly over the respective shapes, i.e.*,

$$\forall \varepsilon > 0 \;\exists\; k_0 \ni k \geq k_0 \Longrightarrow$$

(3.33) $$\frac{K_\lambda(k)}{I(\lambda)} \geq 1 - \varepsilon , \qquad \lambda \in \Lambda ,$$

where

$$K_\lambda(k) = 1'S_\lambda 1$$

and

$$I(\lambda) = I(\theta;\lambda) .$$

Proof: We have seen that the asymptotic variance $\mathrm{Var}[\theta^*_G(k)]$ is of the form $1/n1'S_1 1$, where $\lambda = 1$ is the true value of the shape parameter. We are dealing here with a double limiting process, where first $n \longrightarrow \infty$, then $k \longrightarrow \infty$, but still $k/n \longrightarrow 0$. (This last condition is already implicit in the assumption of the validity of the k-variate asymptotic distribution for every k.)

Thus next, for very large n, we examine $\mathrm{Var}[\theta^*_G(k)]$ for increasing k. In explicit form we have

$$(3.34) \qquad K_\lambda(k) = \sum_{i=o}^{k} \frac{(f_{\lambda,i+1} - f_{\lambda,i})^2}{\lambda_i - \lambda_{i-1}} \quad ,$$

where

$$\lambda_o = 0,\ \lambda_{k+1} = 1,\ f_{\lambda,o} = f_{\lambda,k+1} = 0$$

and

$$f_{\lambda,i} = f(u_i^\lambda|\lambda) \ .$$

It has been shown by Tischendorf (1955) that $K_\lambda(k)$ is a nondecreasing function of k (as additional quantiles are added to a fixed spacing λ_i, $i = 1,\ldots, k$) and that by suitable choice of additional quantiles it can be brought arbitrarily close to its upper bound $E_\lambda[(\partial/\partial u) \log f(u|\lambda)]^2$, i.e.,

$$\forall\ \varepsilon' > 0 \ \exists \ k_\lambda \ni \ k \geq k_\lambda \Longrightarrow$$

$$(3.35) \qquad E_\lambda[\frac{\partial}{\partial u} \log f(u|\lambda)]^2 - K_\lambda(k) < \varepsilon' \ .$$

Now

$$(3.36) \qquad E_\lambda[\frac{\partial}{\partial u} \log f(u|\lambda)]^2 = E_\lambda[\frac{\partial}{\partial \theta} \log f(x - \theta|\lambda)]^2 = I(\lambda)$$

and the efficiency in question is

$$\frac{1}{nI(\lambda)\mathrm{Var}[\theta_G^*(k)]} = \frac{K_\lambda(k)}{I(\lambda)} \ .$$

From (3.35) we have

$$I(\lambda) - K_\lambda(k) < \varepsilon' \ ,$$

so that, since $0 < I(\lambda) < \infty$, $\forall \lambda \in \Lambda$,

$$1 - \frac{K_\lambda(k)}{I(\lambda)} < \frac{\varepsilon'}{I(\lambda)} \leq \varepsilon ,$$

and

$$\frac{K_\lambda(k)}{I(\lambda)} > 1 - \varepsilon . \tag{3.37}$$

Letting $k_o = \max\limits_{\lambda \in \Lambda} (k_\lambda)$, we obtain the result that (3.35) and hence (3.37) hold $\forall k \geq k_o$, as was to be shown. Finally, for any $\varepsilon > 0$, we obtain $\varepsilon'/I(\lambda) \leq \varepsilon$ for each λ by determining ε' above by $\varepsilon' = \varepsilon \min\limits_{\lambda \in \Lambda} (I(\lambda))$.

3.5. Case When Scale Parameter Is Unknown

Unlike the linear estimators to be treated in subsequent chapters, in general the Pitman estimators here discussed are not scale invariant. We have been considering families of shapes of the form $f(x - \theta|\lambda)$, a special case of

$$\frac{1}{\sigma_\lambda} f\left(\frac{x-\theta}{\sigma_\lambda}\Big|\lambda\right) , \tag{3.38}$$

in which σ_λ is known $\forall \; \lambda \; \epsilon \; \Lambda$.

In the case of greater practical interest where σ_λ, as well as λ, is unknown, the situation is more complicated. A systematic generalized Pitman estimator can now be defined by

.39) $\theta_G^{**}(k)$

$$= \sum_{\lambda=1}^{m} \frac{g_\lambda\left(\frac{|S_\lambda|}{1'S_\lambda 1}\right)^{1/2} \exp\{-\frac{n}{2\underline{\hat{\sigma}}_\lambda^2}(\underline{x}^{(k)} - \hat{\sigma}_\lambda \underline{u}^\lambda - \underline{\hat{\theta}}_\lambda)' S_\lambda (\underline{x}^{(k)} - \hat{\sigma}_\lambda \underline{u}^\lambda - \hat{\theta}_\lambda)\}}{\sum_{j=1}^{m} g_j \left(\frac{|S_j|}{1'S_j 1}\right)^{1/2} \exp\{-\frac{n}{2\hat{\sigma}_j^2}(\underline{x}^{(k)} - \hat{\sigma}_j \underline{u}^j - \hat{\theta}_j)' S_j (\underline{x}^{(k)} - \hat{\sigma}_j \underline{u}^j - \hat{\theta}_j)\}} \; \hat{\theta}_\lambda(k) ,$$

where now

$$E_\lambda[X_{(n_i)}] = u_i^\lambda \sigma_\lambda + \theta$$

and $\hat{\sigma}_\lambda$ is an estimator of σ_λ.

Here we need to show, in order to establish results like those of Theorem 3.2, that

$$(3.40)\quad \exp\{-\frac{n}{2\hat{\sigma}_\lambda^2}[(\underline{x}^{(k)}-\hat{\sigma}_\lambda\underline{u}^\lambda-\hat{\theta}_\lambda)'S_\lambda(\underline{x}^{(k)}-\hat{\sigma}_\lambda\underline{u}^\lambda-\hat{\theta}_\lambda)$$

$$-(\underline{x}^{(k)}-\hat{\sigma}_1 u^1-\hat{\theta}_1)'S_1(\underline{x}^{(k)}-\hat{\sigma}_1\underline{u}^1-\hat{\theta}_1)]\} \xrightarrow{P} 0\ ,$$

$\forall\ \lambda \neq 1$, where $\lambda = 1$ is the true value of the shape parameter. Clearly the substitution for each σ_λ of an estimator which is consistent when λ is the true value yields the desired result.

An attractive choice here is the best linear unbiased estimator of σ_λ introduced by Ogawa (1951) and based on the same set of sample quantiles $\forall\ \lambda$.

Using matrix notation, in order to indicate the analogy with (3.18), we obtain

$$(3.41)\qquad \hat{\sigma}_\lambda(k) = \tilde{\sigma}_\lambda(k) = \frac{\underline{u}^\lambda S_\lambda \underline{x}^{(k)}{}'}{\underline{u}^\lambda S_\lambda \underline{u}^\lambda{}'}\quad .$$

Assuming again that $\lambda = 1$ is the true value of the shape parameter, we have

$$(3.42)\qquad E_1[\tilde{\sigma}_1(k)] = \sigma_1 + \theta\,\frac{\underline{u}^1 S_1 1}{\underline{u}^1 S_1 \underline{u}^1{}'}$$

$$= \sigma_1\ ,$$

by comments following (3.25). Also,

$$(3.43)\qquad \mathrm{Var}[\tilde{\sigma}_1(k)|1] = \frac{\underline{u}^1 S_1 \Gamma_1 S_1 \underline{u}^1{}'}{n(\underline{u}^1 S_1 \underline{u}^1{}')^2} \longrightarrow 0 \text{ as } n \longrightarrow \infty,$$

so that

$$\tilde{\sigma}_1(k) \xrightarrow{P} \sigma_1 \quad \text{as} \quad n \longrightarrow \infty .$$

3.6. Discussion

The advantages of using $\theta_G^*(k)$ to estimate location are attractive.

Unlike the other Pitman-type estimators discussed earlier, it is extremely easy to calculate on a computer. Knowledge of the exact covariance matrix of the order statistics involved is not necessary, in contrast to some of the other cases discussed in this paper. $\theta_G^*(k)$ in addition shares the optimum properties of other Pitman estimators; it is admissible within the context of its own model, and approaches full asymptotic efficiency uniformly over all shapes $\lambda \in \Lambda$ even in the context of the exact model and full samples.

It is this latter property that really distinguishes it from its linear competitors, to be discussed in the following chapters.

However, to evaluate its actual performance, numerical work is necessary. For example, how fast are the asymptotic values approached? For small n, how does this k-variate model compare with the full-sample Pitman estimator? How does each of these compare with corresponding linear estimators? Because of the analytic intractability of some of the small-sample variances involved, Monte Carlo studies are being planned for different values of k and $\{\lambda_i\}$, and selected families of underlying distributions.

The problem of the choice of k and $\{\lambda_i\}$ will be treated in Chapter 6.

Chapter 4

SYSTEMATIC LINEAR ESTIMATORS COMPUTED FROM COVARIANCE MATRIX

4.1. Introduction

Using the covariance matrix of the order statistics for a sample of size n, Lloyd (1952) obtained, by means of the generalized Gauss-Markov theorem, best linear unbiased estimators (BLUE's) for the location parameter θ. His method was generalized to the problem of efficiency-robust estimation by Laska (1962), who derived the corresponding admissible linear unbiased estimators (ALUE's).

Lloyd's method may be applied to the exact covariance matrix of a given set of sample quantiles, yielding best systematic linear unbiased estimators (BSLUE's) for given k and spacing $\{\lambda_i\}$. These estimators can then be generalized to provide admissible systematic linear unbiased estimators (ASLUE's) of θ. This work has been carried out in Section 4.2.

If we apply the Gauss-Markov theorem to the covariance matrix of the asymptotic k-variate normal distribution of the sample quantiles, we obtain the estimators studied by Ogawa (1951); we shall call these asymptotically best systematic linear unbiased estimators (ABSLUE's). These estimators are defined in Section 4.3; we then prove, using results of Blom (1958), that the ABSLUE's are asymptotically equivalent to the BSLUE's to within $O_p(n^{-2})$. This is a very useful result, since the exact covariances are available only in a very restricted range of cases. The fact that these ABSLUE's are also asymptotically efficient, as $k \longrightarrow \infty$, in the class of all estimators of θ follows from the work of Tischendorf (1955).

In the last part of Section 4.3 we generalize the ABSLUE's to the problem of efficiency-robust estimation and derive the corresponding asymptotically admissible systematic linear unbiased estimators (AASLUE's); the asymptotic equivalence to the ASLUE's holds analogously, but the asymptotic efficiency does not. That is why we needed in preceding chapters to consider non-linear estimators.

For finite sample sizes it would be desirable to make independent numerical evaluations of the properties of these estimators, in comparison with competing estimators (including those discussed in Chapters 2, 3, and 5) and asymptotic properties.

As before, we assume throughout the chapter that a particular spacing $\{\lambda_i\}$ is given for each k. The problem of optimum spacing will be treated in Chapter 6.

For convenient reference we state here some basic definitions and results. These may be found e.g. in Sarhan and Greenberg (1962) or Blom (1958).

As always, let $X_1, \ldots, X_n$ be n independent, identically distributed random variables with cdf $F(x - \theta)$, and corresponding pdf $f(x - \theta)$, where $E[X_i] = \theta \ \forall\ i$, and $|\theta| < \infty$.

(a) Population quantile

For k given real numbers λ_i, $i = 1, \ldots, k$, $\ni\ 0 < \lambda_1 < \ldots < \lambda_k < 1$, let the λ_i-quantile of the population be x_i, as already stated in (3.7), i.e.,

$$(4.1) \qquad \int_{-\infty}^{x_i} f(t - \theta)dt = \lambda_i , \qquad i = 1, \ldots, k.$$

(b) Sample quantile

Then we may define the k sample quantiles $X_{(n_1)} < X_{(n_2)} < \ldots < X_{(n_k)}$, a subset of the order statistics $X_{(1)} < \ldots < X_{(n)}$, such that

$$n_i = \begin{cases} n\lambda_i, & \text{if } (n\lambda_i) \text{ is an integer} \\ [n\lambda_i]+1 & \text{otherwise,} \quad i = 1, \ldots, k. \end{cases}$$

(c) Exact distribution of sample quantiles

The joint pdf of the k sample quantiles (for $\theta = 0$) is given by

(4.2) $w(\underline{x}^{(k)})$

$$=\begin{cases} \dfrac{n!}{\prod_{i=1}^{k+1}(n_i-n_{i-1}-1)!}\prod_{i=1}^{k+1}[F(x_{(n_i)})-F(x_{(n_{i-1})})]^{n_i-n_{i-1}-1}\prod_{i=1}^{k}f(x_{(n_i)}), \\ \qquad\qquad\qquad\qquad \text{if } x_{(n_1)} < \ldots < x_{(n_k)} \\ 0 \qquad \text{otherwise}, \end{cases}$$

where

$$n_o = 0\ ,\quad n_{k+1} = n+1\ ,\quad x_{(n_o)} = -\infty,\quad x_{(n_{k+1})} = +\infty\ .$$

(d) Exact moments of sample quantiles

(4.3) $E[X_{(n_i)}|\theta = 0]$

$$= v_i = \frac{n!}{(n_i-1)!(n-n_i)!}\int_{-\infty}^{\infty}\xi F^{n_i-1}(\xi)[1-F(\xi)]^{n-n_i}\,dF(\xi)$$

$$= \frac{n!}{(n_i-1)!(n-n_i)!}\int_o^1 G(u)u^{n_i-1}(1-u)^{n-n_i}\,du$$

$$= \int_o^1 G(u)\beta_{n_i}(u;n)du\ ,$$

where $G(u) = F^{-1}(u)$ is assumed to be uniquely determined and $\beta_{n_i}(u;n)$ is the beta function. Second-order moments may be computed from the following:

$$(4.4)\quad E[X_{(n_i)}\, X_{(n_j)} \mid \theta = 0]$$

$$= v_{ij} = \frac{n!}{(n_i-1)!(n_j-n_i-1)!(n-n_j)!} \int_{-\infty}^{\infty} \int_{-\infty}^{\eta} \xi\eta F^{n_i-1}(\xi)$$

$$\cdot\ [F(\eta) - F(\xi)]^{n_j-n_i-1} [1 - F(\eta)]^{n-n_j} dF(\xi)dF(\eta)$$

$$= \iint\limits_{o<u<v<1} G(u)G(v)\beta_{n_i,n_j}(u,v;n)dudv\ ,$$

where

$$\beta_{ij}(u,v;n) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} u^{i-1}(v-u)^{j-i-1}(1-v)^{n-j}\ ,$$

$$i < j\ .$$

We obtain, for the variance,

$$(4.5)\quad \mathrm{Var}[X_{(n_i)}] = \int_o^1 [G(u)]^2\, \beta_{n_i}(u;n)du - \left[\int_o^1 G(u)\beta_{n_i}(u;n)du\right]^2$$

These integrals do not always exist.

(e) Sufficient conditions for existence of moments of sample quantiles

Blom (1958), p. 43, gives two sets of sufficient conditions; we will state only the stronger one (since it is simpler and sufficient for our purposes):

$$|E[X_{(n_i)}]| < \infty \quad \forall\, i \quad \text{if} \quad E|X| < \infty$$

and

$$|E[X_{(n_i)}\, X_{(n_j)}]| < \infty \quad \forall\ i,j \quad \text{if} \quad EX^2 < \infty .$$

(Of course, this holds in our case, since we have $|\theta| < \infty$, $\sigma = 1$).

(f) Asymptotic distribution of sample quantiles

This is Mosteller's theorem which was stated already in full detail as Theorem 3.1.

(g) Limiting form of first and second-order moments of full set of order statistics

These results are a slightly modified form of Blom's (1958), p. 53ff. Unlike him, we assume the stronger set of sufficient conditions for the existence of the exact moments.

First we need the following definitions.

DEFINITION. c-sequence. The double sequence $\{x_{i_\nu, n_\nu}\}$ $\nu = 1, 2, 3, \ldots,\ n_\nu < n_{\nu+1}$, is called a c-sequence if

$$(4.6) \qquad \lim_{\nu \to \infty} \frac{i_\nu}{n_\nu} = c \ , \qquad 0 \leq c \leq 1 \ .$$

DEFINITION. Less than $n^{-1/2}$-convergent c-sequence. $\{x_{i_\nu, n_\nu}\}$ is a less than $n^{-1/2}$-convergent c-sequence if $\exists$ M, $\rho \ni$ for all but a finite number of ν's

$$(4.7) \qquad \left| \frac{i_\nu}{n_\nu} - c \right| > M n_\nu^{-1/2+\rho} \qquad .$$

Then we have,

LEMMA 4.1. (Mean) *If*

(a) $E|X| < \infty$

(b) $G(u)$ *and* $G'(u)$ *are bounded and continuous, and* $G''(u)$ *is bounded in* $\mathcal{C} = [0,1]$, *with the possible exception of a finite number of points* $\mathcal{C}^*$

(c) X_i *belongs to any c-sequence when* $c \in \mathcal{C}$, *and to a less than* $n^{-1/2}$*-convergent c-sequence when* $c \in \mathcal{C}^*$,

then

$$E[X_i|\theta = 0] = G(\tfrac{i}{n+1}) + R_i , \tag{4.8}$$

where $|R_i| < \frac{M}{n}$, *and* M *is independent of* i *and* n.

LEMMA 4.2. (Second-order moments) *If*

(a) $EX^2 < \infty$

(b) $G(u)$, $G'(u)$ *and* $G''(u)$ *are bounded and continuous and* $G'''(u)$ *is bounded in* $\mathcal{C} = [0,1]$ *with the possible exception of a finite number of points* $\mathcal{C}^*$,

(c) X_i *belongs to any* c_1*-sequence when* $c_1 \in \mathcal{C}$, *and to a less than* $n^{-1/2}$*-convergent* c_1*-sequence when* $c_1 \in \mathcal{C}^*$; X_j *belongs to a* c_2*-sequence* $(c_1 \leq c_2)$ *when* $c_2 \in \mathcal{C}$, *and to a less than* $n^{-1/2}$*-convergent* c_2*-sequence when* $c_2 \in \mathcal{C}^*$,

then

$$\operatorname{Cov}(X_i,X_j) = \frac{1}{n} p_i(1-p_j)G'(p_i)G'(p_j) + R_{ij} , \qquad i \leq j, \tag{4.9}$$

where

$$p_i = \frac{i}{n+1}\ , \qquad p_j = \frac{j}{n+1}$$

and

$$|R_{ij}| < \frac{M}{n^2}\ ,$$

with M independent of i, j, and n.

(These lemmas will be used also in Section 5.1. For convenience of application we have changed the factor $(n+2)^{-1}$ in (4.9) to n^{-1}; the difference is of course absorbed in R_{ij}.)

(h) Application of (g) to k sample quantiles

Modifying the above results to the case of k sample quantiles we may write, since

$$(4.10) \qquad \lim_{k\to\infty}\ \lim_{n\to\infty} \frac{n_i}{n+1} = \lambda_i\ , \qquad i = 1, \ldots, k\ ,$$

and G(u) is continuous,

$$(4.11) \qquad E[X_{(n_i)} | \theta = 0] = G(\lambda_i) + O(n^{-1})$$

and

$$(4.12) \qquad \mathrm{Cov}[X_{(n_i)}, X_{(n_j)}] = \frac{1}{n}\lambda_i(1-\lambda_j)G'(\lambda_i)G'(\lambda_j) + O(n^{-2})\ .$$

Since $G'(u) = 1/f(x)$, we see that this application of Lemmas 4.1 and 4.2 provides sufficient conditions for the given moments to converge to the corresponding moments of the limiting distribution (cf. (3.10)).

In (3.8) we defined u_i as the λ_i-quantile of the standardized distribution,

$$u_i = F^{-1}(\lambda_i) = G(\lambda_i) \ . \tag{4.13}$$

Indicating a non-standardized distribution by underlining, we have

$$\int_{-\infty}^{x_i} \underline{f}(y)dy = \lambda_i \ , \tag{4.14}$$

so that, by (4.11),

$$\begin{aligned} E[X_{(n_i)}] &= \underline{F}^{-1}(\lambda_i) + O(\tfrac{1}{n}) \\ &= x_i + O(\tfrac{1}{n}) \ . \end{aligned} \tag{4.15}$$

But since

$$\underline{f}(y) = f(t - \theta) \ , \tag{4.16}$$

we obtain

$$\int_{-\infty}^{x_i} f(t - \theta)dt = F(x_i - \theta) = \lambda_i \ , \tag{4.17}$$

so that

$$\underline{G}(\lambda_i) = x_i = G(\lambda_i) + \theta = u_i + \theta \ , \tag{4.18}$$

yielding, by (4.15),

$$E[X_{(n_i)}] = G(\lambda) + \theta + O(\tfrac{1}{n}) \ . \tag{4.19}$$

4.2. Systematic Linear Estimators Based on Exact Covariance Matrix

In this section we generalize to the case of systematic estimators of θ the results of Lloyd (1952) and Laska (1962); (C1) is assumed to hold.

THEOREM 4.1. (BSLUE's) *Let* $X_{(n_1)}, \ldots, X_{(n_k)}$ *be a set of sample quantiles such that*

$$E[X_{(n_i)}] = v_i + \theta \ , \tag{4.20}$$

and

$$\mathrm{Cov}[X_{(n_i)}, X_{(n_j)}] = t_{ij} = v_{ij} - v_i v_j \ , \tag{4.21}$$

where v_{ij}, v_i, v_j *are defined by* (4.3) *and* (4.4). *Let*

$$T = [t_{ij}] \ , \qquad i, j = 1, \ldots, k \ ,$$

such that T^{-1} *exists, and let*

$$1 = (1, \ldots, 1)' \ , \quad \text{a } k\text{-vector} \ .$$

Then

$$\tilde{\theta}(k,n) = \frac{1'T^{-1}\underline{x}^{(k)}{}'}{1'T^{-1}1} \tag{4.22}$$

is a best systematic linear unbiased estimator (*BSLUE*) *of* θ.

Proof: The theorem follows immediately from application of the generalized Gauss-Markov theorem to T, for estimating the location parameter when the underlying distribution is symmetric, as shown by Lloyd (1952). Clearly $\tilde{\theta}(k,n)$ has minimum variance among linear unbiased estimators based on the given k sample quantiles, the model now being considered.

THEOREM 4.2. (ASLUE's) *Assume that the conditions of Theorem 4.1 hold for each of a family of* m *shapes* λ, $\lambda = 1, \ldots, m$, *where now*

$$E_\lambda[X_{(n_i)}] = v_i^\lambda + \theta \tag{4.23}$$

and

$$\mathrm{Cov}_\lambda[X_{(n_i)}, X_{(n_j)}] = [t_{ij}^\lambda] = T_\lambda \ . \tag{4.24}$$

Further, let

$$M = \sum_{\lambda=1}^{m} g_\lambda T_\lambda \ , \tag{4.25}$$

where $G = \{g_1, \ldots, g_m\}$ *is a given set of real numbers* $\ni \sum_{\lambda=1}^{m} g_\lambda = 1$ *and* $0 < g_\lambda < 1 \ \forall \ \lambda$. *From* (4.25) *it follows that* M *is positive definite and hence non-singular. Then*

$$\tilde{\theta}_G(k,n) = \frac{1'M^{-1}\underline{x}^{(k)}}{1'M^{-1}1} \tag{4.26}$$

is an admissible systematic linear unbiased estimator (ASLUE) of θ.

Proof: The method of proof is completely analogous to that of Laska (1962) for the case of a full sample. Admissibility here is claimed over $\Omega = \Theta \times \Lambda$ (with variance as the risk function), in the class of linear unbiased estimators based on the k sample quantiles.

The proof involves applying the Gauss-Markov theorem to the mixture model (2.8); this yields the estimator

$$\hat{\theta}_G(k,n) = \frac{1'T_G^{-1}\underline{x}^{(k)'}}{1'T_G^{-1}1} , \tag{4.27}$$

where T_G is the covariance matrix of $\underline{x}^{(k)}$ under the mixture model. Although in general

$$T_G \neq \sum_{\lambda=1}^{m} g_\lambda T_\lambda = M ,$$

it has been shown by Craig (1967) that under the symmetry assumptions on $f(x|\theta)$ and the λ_i (defined by (3.2)), valid in this case,

$$1'T_G^{-1}1 = 1'M^{-1}1 \tag{4.28}$$

and

$$1'T_G^{-1}\underline{x}^{(k)'} = 1'M^{-1}\underline{x}^{(k)'} , \tag{4.29}$$

so that

$$\hat{\theta}_G(k,n) = \tilde{\theta}_G(k,n) .$$

This has great computational significance, since we need consider only M. (Cf. also the discussion in Birnbaum and Laska (1967a).)

Admissibility is now easy to show by the usual method: Assume $\tilde{\theta}_G(k,n)$ is inadmissible; then $\exists$ a linear estimator $\theta'(k,n)$ such that

$$\text{(4.30)} \qquad \mathrm{Var}[\theta'|\lambda] \leq \mathrm{Var}[\tilde{\theta}|\lambda] \ , \quad \forall\ \lambda \in \Lambda \ ,$$

with inequality strict for at least one $\lambda \in \Lambda$.

As linear estimators, both θ' and $\tilde{\theta}$ have variances independent of θ; in particular,

$$\text{(4.31)} \qquad \mathrm{Var}[\tilde{\theta}_G|G] = \frac{1}{n1'M^{-1}1} \ .$$

Now

$$\begin{aligned} \mathrm{Var}[\theta'|G] &= \sum_{\lambda=1}^{m} \mathrm{Var}[\theta'|\lambda] && \text{by (2.9)} \\ &< \sum_{\lambda=1}^{m} \mathrm{Var}[\tilde{\theta}|\lambda] && \text{by (4.30)} \\ &= \mathrm{Var}[\tilde{\theta}|G] && \text{by (2.9),} \end{aligned}$$

but this contradicts the assumption that $\tilde{\theta}_G$ is, by the Gauss-Markov theorem, the unique BLUE of θ under the mixture model considered.

4.3. Systematic Linear Estimators Based on Asymptotic Covariance Matrix

We now define Ogawa's estimator of θ and show that it is asymptotically (as $n \longrightarrow \infty$) equivalent to the best systematic linear unbiased estimator $\tilde{\theta}(k,n)$. By the result of Tischendorf described in the proof of Theorem 3.3, such estimators also approach asymptotic efficiency as we let $k \longrightarrow \infty$. As always $f(x - \theta)$ is assumed to be symmetric.

THEOREM 4.3. (ABSLUE's) *Let* $X_{(n_1)}, \ldots, X_{(n_k)}$ *be a set of sample quantiles with asymptotic moments*

$$E[X_{(n_i)}] = u_i + \theta \tag{4.32}$$

and

$$\mathrm{Cov}[X_{(n_i)}, X_{(n_j)}] = \frac{1}{n}\Gamma = \frac{\lambda_i(1-\lambda_j)}{nf(u_i)f(u_j)}$$

as defined by (4.19) and (4.12); we are disregarding the remainder terms. Let $S = \Gamma^{-1}$ *exist, and let* $1 = (1,\ldots,1)'$. *Then*

$$\tilde{\theta}(k) = \frac{1'S\underline{x}^{(k)'}}{1'S1} \tag{4.33}$$

is an asymptotically best systematic linear unbiased estimator (ABSLUE) of θ; *that is,*

$$\tilde{\theta}(k,n) = \tilde{\theta}(k) + O_p(n^{-2}) , \tag{4.34}$$

where $\tilde{\theta}(k,n)$ *is as defined by (4.22).*

Proof: That $\tilde{\theta}(k)$ is best among linear unbiased estimators based on the asymptotic k-variate model, follows immediately from application of the generalized Gauss-Markov theorem.

It is similarly easy to show that it is asymptotically, as $n \longrightarrow \infty$, equivalent to the BSLUE based on the exact model, as indicated by (4.34).

We have

$$\tilde{\theta}(k,n) = \frac{1'T^{-1}\underline{x}^{(k)},}{1'T^{-1}1} , \tag{4.35}$$

where, by (4.12),

$$t_{ij} = \frac{\lambda_i(1-\lambda_j)}{nf(u_i)f(u_j)} + O(n^{-2}) , \tag{4.36}$$

or

$$T = \frac{1}{n} \Gamma + [O(n^{-2})] , \tag{4.37}$$

where the last term represents a k × k matrix each of whose elements is of the form $O(n^{-2})$.

We now substitute (4.37) in the expression for T in (4.35) and carry out all operations as indicated, using standard rules for working with $O(\cdot)$.

From the inequality

$$P\{|X| \geq \varepsilon\} \leq \frac{E|X|}{\varepsilon} \tag{4.38}$$

(Markov's Inequality for $r = 1$) we can derive the relationship

$$X_{(n_i)} = O_p(E|X_{n_i}|) \; ,$$

and finally, in simplest form,

(4.39) $$\sum_{i=1}^{k} X_{(n_i)} = O_p(1) \; .$$

Application of (4.39), together with the standard operations indicated above, yields the desired result as stated in (4.34).

THEOREM 4.4. $\tilde{\theta}(k)$ _as defined in_ (4.33) _is asymptotically efficient as_ k ⟶ ∞, _in the class of all unbiased estimators of_ θ.

Proof: This result was proved by Tischendorf (1955); cf. discussion above in the proof of Theorem 3.3.

Theorem 4.3 may be generalized to the problem of efficiency-robust estimation. This will be done in the following

THEOREM 4.5. (AASLUE's) _Assume that the conditions of Theorem_ 4.3 _hold for each of a family of_ m _shapes_ λ, λ = 1,..., m, _where now_

(4.40) $$E_\lambda[X_{(n_i)}] = u_i^\lambda + \theta$$

and

$$\mathrm{Cov}_\lambda[X_{(n_i)}, X_{(n_j)}] = \frac{1}{n}\Gamma_\lambda = \frac{\lambda_i(1-\lambda_j)}{nf(u_i^\lambda|\lambda)f(u_j^\lambda|\lambda)} \; .$$

Further let

$$N = \sum_{\lambda=1}^{m} g_\lambda \Gamma_\lambda .$$

Then

$$\tilde{\theta}_G(k) = \frac{1'N^{-1}\underline{x}^{(k)}}{1'N^{-1}1} \tag{4.41}$$

is an asymptotically admissible systematic linear unbiased estimator (AASLUE) of θ; specifically,

$$\tilde{\theta}_G(k,n) = \tilde{\theta}_G(k) + O_p(n^{-2}) . \tag{4.42}$$

Proof: The arguments are completely analogous to those given in the proofs of Theorems 4.2 and 4.3.

4.4. Discussion

As is seen from the definition of the respective estimators, knowledge of the exact covariance matrix of the sample quantiles for the pdf's under consideration is needed for the computation of BSLUE's and ASLUE's. A list of available moments, other related numerical results, and a more extensive discussion of this problem-area are found in Birnbaum and Laska (1967a, b).

The estimators based on the asymptotic covariance matrix are more practically applicable, since the asymptotic form is easily computed for arbitrary given pdf's. The necessary inversion of the matrix in question does not appear prohibitive, since here we are dealing only with a subsample of size k, and there are indications that small values of k already yield good results. This is an important practical advantage of this type of systematic estimator.

As shown in Theorems 4.3 and 4.5, the ABSLUE's and AASLUE's are also asymptotically equivalent to the corresponding exact-covariance estimators. But, full asymptotic efficiency, attainable by single-shape estimators, is not attainable for the mixture model even when the full sample and exact covariances are used. This of course was the reason for introducing above the non-linear estimators of the generalized Pitman type.

For a single shape and the full sample an asymptotic form of $\tilde{\theta}(n,n)$ was obtained by Daniell (1920), Bennett (1952), and Jung (1955). The problem of deriving the analogous asymptotic form for the mixture model by Bennett's method seems formidable, as it would require an explicit expression for $[\sum_{\lambda=1}^{m} g_{\lambda}\Gamma_{\lambda}]^{-1}$. However, Jung's approach lends itself to adaptation, as Yhap (1966, 1967) has shown. In the next chapter we will discuss the latter approach as it applies to systematic estimation.

The need for numerical investigation of these estimators has already been mentioned in earlier sections.

Chapter 5

SYSTEMATIC LINEAR ESTIMATORS BASED ON CONTINUOUS WEIGHT FUNCTIONS

5.1. Derivation of Jung's Result for a Single Shape Under Weakened (Necessary and Sufficient) Conditions

As before, let $X_1, \ldots, X_n$ be a set of independent, identically distributed random variables with pdf $f(x - \theta)$ satisfying (C1).

In the last section we mentioned the asymptotic version of Lloyd's best linear estimators obtained by Bennett (1952), Jung (1955), and (less rigorously) by Daniell (1920),

$$\overline{\theta}(h) = \frac{1}{n} \sum_{i=1}^{n} h\left(\frac{i}{n+1}\right) X_{(i)} , \tag{5.1}$$

where $X_{(1)} < \ldots < X_{(n)}$ are the order statistics derived from $X_1, \ldots, X_n$, and $h(u)$ is a continuous function of u defined on $[0,1]$. It was also shown by these authors that $\overline{\theta}(h)$ is asymptotically efficient in the class of all estimators of θ, under the given shape.

In this section we will derive Jung's result under weakened conditions and indicate in what sense these conditions are necessary and sufficient.

Let

$$Z_n = \frac{1}{n} \sum_{i=1}^{n} h\left(\frac{i}{n+1}\right) Y_{(i)} , \tag{5.2}$$

where

$$(5.3) \qquad E[Y_{(i)}] = E[X_{(i)}] - \theta = E[X_{(i)}|\theta = 0] ,$$

i.e., we are considering first the estimator based on standardized variables.

LEMMA 5.1. Let h(u) be continuous and bounded on [0,1]. Assume further that G(u), G'(u), and G"(u) are bounded and continuous and G'"(u) is bounded on [0,1], where $G(u) = F^{-1}(u) = y$. Then

$$(5.4) \qquad E[Z_n] = \int_o^1 h(u)G(u)du + O(n^{-1})$$

and

$$(5.5) \qquad Var[Z_n] = \frac{1}{n}\int_o^1 \int_o^1 k(u,v)h(u)h(v)G'(u)G'(v)dudv + O(n^{-2}),$$

where

$$(5.6) \qquad k(u,v) = \begin{cases} u(1-v) & u \leq v \\ v(1-u) & u > v . \end{cases}$$

Proof: We have

$$(5.7) \qquad E[Z_n] = \frac{1}{n}\sum_{i=1}^{n} h(\tfrac{i}{n+1})E[Y_{(i)}] .$$

Since the assumptions of Lemma 4.1 are fulfilled in this case, we have, by (4.8),

$$(5.8)\quad E[Z_n] = \frac{1}{n}\sum_{i=1}^{n} h(\tfrac{i}{n+1})[G(\tfrac{i}{n+1}) + O(n^{-1})]$$

$$= \frac{1}{n}\sum_{i=1}^{n} h(\tfrac{i}{n+1})G(\tfrac{i}{n+1}) + \frac{1}{n}\sum_{i=1}^{n} h(\tfrac{i}{n+1}) \cdot O(n^{-1}) .$$

Let

$$\frac{1}{n} = \Delta y_i , \qquad \frac{i}{n+1} = \xi_i , \qquad \frac{i}{n} = y_i , \qquad i = 1,\ldots, n;$$

then

$$y_{i-1} < \xi_i < y_i ,$$

and since h(u) and G(u) are continuous and bounded on [0,1], the Riemann integral exists, so that in the limit

$$(5.9)\qquad E[Z_n] = \int_0^1 h(u)G(u)du + O(n^{-1})$$

since also

$$(5.10)\qquad \frac{1}{n}\sum_{i=1}^{n} h(\tfrac{i}{n+1}) = O(1) \quad \forall\, n .$$

Similarly,

$$(5.11)\qquad Var[Z_n] = \frac{1}{n^2}\, h'(n)\ Cov[Y_{(i)},Y_{(j)}]h(n) ,$$

where

$$h(n) = [h(\tfrac{1}{n+1}),\ldots, h(\tfrac{n}{n+1})]' ,$$

so that

$$(5.12)\qquad \mathrm{Var}[Z_n] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} h(\tfrac{i}{n+1})h(\tfrac{j}{n+1})\mathrm{Cov}[Y_{(i)},Y_{(j)}]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} h(\tfrac{i}{n+1})h(\tfrac{j}{n+1})\Big[\frac{1}{n}\, k(\tfrac{i}{n+1}, \tfrac{j}{n+1})G'(\tfrac{i}{n+1})$$

$$\cdot\, G(\tfrac{j}{n+1}) + O(n^{-2})\Big] ,$$

where

$$(5.13)\qquad k(\tfrac{i}{n+1}, \tfrac{j}{n+1}) = \begin{cases} \frac{i}{n+1}\,(1 - \frac{j}{n+1}), & i \le j \\ \frac{j}{n+1}\,(1 - \frac{i}{n+1}), & i > j\ ; \end{cases}$$

we have here used (4.9) of Lemma 4.2, which applies by virtue of our assumptions. By the same arguments as before we establish the existence of the double Riemann integral; that, with (5.10), yields in the limit

$$\mathrm{Var}[Z_n] = \frac{1}{n}\int_0^1\int_0^1 k(u,v)h(u)h(v)G'(u)G'(v)\,du\,dv + O(n^{-2}) ,$$

as was to be shown.

COROLLARY 5.1. Let

$$(5.14)\qquad \bar{\theta}(h) = \frac{1}{n}\sum_{i=1}^{n} h(\tfrac{i}{n+1})X_{(i)}\ ;$$

then

$$(5.15)\qquad E[\bar{\theta}(h)] = \int_0^1 h(u)G(u)\,du + \theta\int_0^1 h(u)\,du + O(n^{-1})$$

and

$$(5.16)\quad \mathrm{Var}[\overline{\theta}(h)] = \frac{1}{n}\int_0^1\int_0^1 k(u,v)h(u)h(v)G'(u)G'(v)\,du\,dv + O(n^{-2}) .$$

Proof: (5.15) follows from (5.2) - (5.4) and (5.16) is an immediate consequence of (5.5).

We will now derive Jung's result for the special case of estimating θ, with $f(x - \theta)$ symmetric, using the results of Corollary 5.1.

The proof is essentially that of Jung, based on the calculus of variations. However, we give it in a somewhat different form which will make it more convenient to apply in later sections; the terminology follows that of Yhap (1967).

THEOREM 5.1. If the assumptions of Lemma 5.1 hold and

$$(5.17)\quad |D| = \begin{vmatrix} \int_0^1 h_o(v)\,dv & \int_0^1 h_1(v)\,dv \\ \int_0^1 h_o(v)G(v)\,dv & \int_0^1 h_1(v)G(v)\,dv \end{vmatrix} \neq 0 ,$$

where

$$(5.18)\quad h_o(v) = \frac{1}{G'(v)}\frac{d^2}{dv^2}\left[\frac{1}{G'(v)}\right]$$

and

$$h_1(v) = \frac{1}{G'(v)} \frac{d^2}{dv^2} \left[\frac{G(v)}{G'(v)}\right] ,$$

then there is, in the class of continuous weight functions h(u), *one that uniquely minimizes the asymptotic variance* (5.16), *namely,*

$$h(u) = \frac{h_o(u)}{\int_o^1 h_o(u)du} . \tag{5.19}$$

Proof: Lemma 5.1 guarantees the existence of the asymptotic mean and variance of $\overline{\theta}(h)$ as given by (5.15) and (5.16). We want to find h minimizing the functional

$$n \operatorname{Var}[\overline{\theta}(h)] = \int_o^1 \int_o^1 k(u,v)G'(u)G'(v)h(u)h(v)dudv$$

subject to the subsidiary conditions

$$\int_o^1 h(u)du = 1 \quad \text{and} \quad \int_o^1 h(u)G(u)du = 0 . \tag{5.20}$$

From (5.15) we see that these are necessary to insure the (asymptotic) unbiasedness of $\overline{\theta}(h)$.

Writing $n \operatorname{Var}[\overline{\theta}(h)]$ as

$$\int_o^1 K(v,h(v))dv = \int_o^1 G'(v)h(v)\left[\int_o^1 k(u,v)G'(u)h(u)du\right]dv , \tag{5.21}$$

we obtain for the Euler equation

$$[K]_h = G'(v) \int_o^1 k(u,v)G'(u)h(u)du + \lambda_o + \lambda_1 G(v) = 0 , \tag{5.22}$$

where λ_o, λ_1 are the Lagrange multipliers. Solving for h, we have

$$(5.23) \qquad \int_o^1 k(u,v)G'(u)h(u)du = \frac{-\lambda_o}{G'(v)} - \frac{\lambda_1 G(v)}{G'(v)} ,$$

or

$$(5.24) \quad (1-v)\int_o^v uG'(u)h(u)du + v\int_v^1 (1-u)G'(u)h(u)$$

$$= \frac{-\lambda_o}{G'(v)} - \frac{\lambda_1 G(v)}{G'(v)} .$$

Differentiation with respect to v yields

$$(5.25) \quad -\int_o^v uG'(u)h(u)du + (1-v)vG'(v)h(v) + \int_v^1 (1-u)G'(u)h(u)du$$

$$- v(1-v)G'(v)h(v) = -\lambda_o \frac{d}{dv}\left[\frac{1}{G'(v)}\right] - \lambda_1 \frac{d}{dv}\left[\frac{G(v)}{G'(v)}\right] .$$

Differentiating again, we obtain

$$(5.26) \qquad -vG'(v)h(v) - (1-v)G'(v)h(v)$$

$$= -\lambda_o \frac{d^2}{dv^2}\left[\frac{1}{G'(v)}\right] - \lambda_1 \frac{d^2}{dv^2}\left[\frac{G(v)}{G'(v)}\right]$$

so that

$$(5.27) \quad h(v) = \frac{\lambda_o}{G'(v)} \frac{d^2}{dv^2}\left[\frac{1}{G'(v)}\right] + \frac{\lambda_1}{G'(v)} \frac{d^2}{dv^2}\left[\frac{G(v)}{G'(v)}\right]$$

$$= \lambda_o h_o(v) + \lambda_1 h_1(v) .$$

To solve for λ_o and λ_1, we need the constraints (5.20). Substituting, we have

(5.28) $$\lambda_o \int_o^1 h_o(v)dv + \lambda_1 \int_o^1 h_1(v)dv = 1$$

and

(5.29) $$\lambda_o \int_o^1 h_o(v)G(v)dv + \lambda_1 \int_o^1 h_1(v)G(v)dv = 0 .$$

In matrix form, this may be written as

(5.30) $$\begin{vmatrix} \int_o^1 h_o(v)dv & \int_o^1 h_1(v)dv \\ \int_o^1 h_o(v)G(v)dv & \int_o^1 h_1(v)G(v)dv \end{vmatrix} \begin{vmatrix} \lambda_o \\ \lambda_1 \end{vmatrix} = \begin{vmatrix} 1 \\ 0 \end{vmatrix}$$

i.e.,

$$D\lambda = \alpha ,$$

so that

(5.31) $$\lambda = D^{-1}\alpha ,$$

since by (5.17) $|D| \neq 0$.

It has been shown by Yhap (1967) that for f(x) symmetric,

(5.32) $$\int_o^1 h_o(v)G(v)dv = 0 .$$

Solving for λ, we thus obtain

$$\lambda_o = \frac{1}{\int_o^1 h_o(v)dv} ,$$

and

$$\lambda_1 = 0 ,$$

which yields the desired result (5.19).

Uniqueness of the solution has been shown by Jung and the argument will not be given here, as it is not pertinent to our present considerations.

It has likewise been shown by Jung that $\mathrm{Var}[\overline{\theta}(h)]$, when h is the minimizing function h(u) defined by (5.19), is the reciprocal of Fisher's Information for θ, so that $\overline{\theta}(h)$ is asymptotically efficient in the class of all estimators. Asymptotic normality of this estimator, required for efficiency, was shown by Chernoff, Gastwirth, and Johns (1967).

We see from the above that sufficient conditions for Theorem 5.1 consist essentially of $G'''(u)$ being bounded and h(u) integrable on [0,1], these assuring the existence of the asymptotic moments in question. Inspection of (5.19) and (5.18) tells us that these conditions are also necessary for the minimizing function h(u) to exist, and this is precisely the extent of our new results.

Jung, in establishing his analogue of Lemma 5.1, required the existence of a bounded fourth derivative on h(u). In view of the form of h(u) as given in (5.19), his condition amounts to the requirement of seven bounded derivatives on G(u), as compared to our three.

In his concluding remarks Jung pointed to the desirability of further weakening these conditions, allowing h(u) to tend to ∞ at $u = 0$ or $u = 1$, provided that remainder terms for the asymptotic moments could still be obtained that approached 0 not too slowly. Now in our results we made use of only the weakest form of Lemmas 4.1 and 4.2, assuming condition (b) to hold for all of $\mathcal{C} = [0,1]$. It is possible that investigation of the full use of these lemmas would lead to results along the lines indicated by Jung.

5.2. Direct Analogue for Systematic Estimation and Arbitrary Spacing

In this section we apply the method of systematic estimation to the case discussed in Section 5.1, deriving asymptotically best systematic linear unbiased estimators. These estimators are valid for any k and arbitrary spacing $\{\lambda_i\}$; for the uniform spacing $\lambda_i = i/(k+1)$, $i = 1,\dots, k$, the weight function h(u) is seen to coincide with the one given by Jung for the full sample.

These results were obtained already by Särndal (1962) by a different method (extending the approach of Bennett and Ogawa). Our work is of interest here in that it yields the same results in a simpler and more direct manner, and by the only approach that seems to be adaptable to the problem of efficiency-robust estimation.

In what follows we will make use of a generating function $f^*(x)$ for $\{\lambda_i\}$ introduced by Särndal (1962), which will also be of central interest in the next chapter. Our definition of f*(x) will be specialized to the case of interest in this paper.

DEFINITION. Let f*(x) be a function continuous, bounded, and positive $\forall x$, and $\ni \int_{-\infty}^{\infty} f^*(x)dx = 1$. Assume further that f*'(x) is continuous and bounded $\forall x$. Let

$$F^*(x) = \int_{-\infty}^{x} f^*(t)dt , \tag{5.33}$$

$$G^*(t) = (F^*)^{-1}(t) , \tag{5.34}$$

where G*(t) is continuous and nondecreasing, and $\int_{0}^{1} G^*(t)dt < \infty$. Finally, let

$$\pi_i = \frac{i}{k+1} , \quad i = 1, \ldots, k . \tag{5.35}$$

The set of spacings $\{\lambda_i\}$ is then defined in terms of f*(x) as follows:

$$\lambda_i = F[G^*(\pi_i)] , \quad i = 1, \ldots, k , \tag{5.36}$$

where F(x) is the cdf of the random variable X. Clearly $\lambda_i < \lambda_{i+1}$, $i = 1, \ldots, k-1$. We shall say that the spacing (5.36) is generated by f*(x), the "generating function" of $\{\lambda_i\}$.

LEMMA 5.2. Given a set of k sample quantiles $X_{(n_1)}, \ldots, X_{(n_k)}$ determined by $\{\lambda_i\}$, and a function h(u) continuous and bounded on [0,1], define the systematic linear estimator

$$Z_k = \frac{1}{k}\sum_{i=1}^{k} h(\lambda_i)Y_{(n_i)} , \tag{5.37}$$

where

$$E[Y_{(n_i)}] = E[X_{(n_i)}] - \theta = E[X_{(n_i)}|\theta = 0],$$

$$i = 1,\ldots, k .$$

Assume further that G(u), G'(u), and G"(u) are bounded and continuous and G'"(u) is bounded on [0,1]. Then

$$E[Z_k] = \int_0^1 h(u)G(u)G'(u)f^*[G(u)]du + O(n^{-1}) \tag{5.38}$$

and

$$\mathrm{Var}[Z_k] = \frac{1}{n}\int_0^1\int_0^1 k(u,v)h(u)h(v)[G'(u)]^2 \cdot [G'(v)]^2 f^*[G(u)]f^*[G(v)]dudv + O(n^{-2}) , \tag{5.39}$$

where f*(x) is defined by (5.36), and k(u,v) by (5.6).

Proof: The proof is quite similar to that of Lemma 5.1. In view of the assumptions on G(u), (4.11) and (4.12) are valid, so that we obtain

$$(5.40)\qquad E[Z_k] = \frac{1}{k}\sum_{i=1}^{k} h(\lambda_i)E[Y_{(n_i)}]$$

$$= \frac{1}{k}\sum_{i=1}^{k} h(\lambda_i)[G(\lambda_i) + O(n^{-1})]$$

$$= \frac{1}{k}\sum_{i=1}^{k} h(\lambda_i)G(\lambda_i) + O(n^{-1})$$

$$= \frac{1}{k}\sum_{i=1}^{k} h[F[G^*(\tfrac{i}{k+1})]]G^*(\tfrac{i}{k+1}) + O(n^{-1}) \text{ .}$$

Let

$$\frac{1}{k} = \Delta y_i \text{ ,} \qquad \frac{i}{k+1} = \xi_i \text{ ,} \qquad \frac{i}{k} = y_i \text{ ,} \qquad i = 1,\ldots, k \text{ ;}$$

then

$$y_{i-1} < \xi_i < y_i \text{ ,}$$

and by the assumptions on $h(\cdot)$ and $G^*(\cdot)$ the Riemann integral exists, so that we obtain in the limit

$$(5.41)\qquad E[Z_k] = \int_0^1 h[F[G^*(v)]]G^*(v)dv + O(n^{-1}) \text{ .}$$

Now let

$$F[G^*(v)] = u$$

$$G^*(v) = G(u) = x$$

and

$$\frac{dv}{du} = \frac{dF^*}{dG} \cdot \frac{dG}{du} = \frac{f^*(x)}{f(x)} = G'(u)f^*[G(u)] \text{ ,}$$

so that (5.41) becomes

$$(5.42)\quad E[Z_k] = \int_o^1 h(u)G(u)G'(u)f^*[G(u)]du + O(n^{-1}).$$

Similarly, by (4.12),

$$(5.43)\quad \mathrm{Var}[Z_k] = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k} h(\lambda_i)h(\lambda_j)\mathrm{Cov}[Y_{(n_i)},Y_{(n_j)}]$$

$$= \frac{1}{nk^2}\sum_{i=1}^{k}\sum_{j=1}^{k} h(\lambda_i)h(\lambda_j)[k(\lambda_i,\lambda_j)G'(\lambda_i)G'(\lambda_j)+O(n^{-2})],$$

where

$$(5.44)\qquad k(\lambda_i,\lambda_j)=\begin{cases}\lambda_i(1-\lambda_j) & i \leq j\\ \\ \lambda_j(1-\lambda_i) & i > j\ .\end{cases}$$

Substituting (5.36) for λ_i as before, and by similar arguments for the existence of the double Riemann integral, we obtain in the limit as $k \longrightarrow \infty$

$$(5.45)\quad \mathrm{Var}[Z_k] = \frac{1}{n}\int_o^1\int_o^1 h[F[G^*(s)]]h[F[G^*(t)]]$$

$$\cdot\ k[F[G^*(s)],F[G^*(t)]]G'[F[G^*(s)]]G'[F[G^*(t)]]dsdt + O(n^{-2}).$$

Performing the change of variables

$$F[G^*(s)] = u$$

and

$$F[G^*(t)] = v$$

now yields the desired result.

COROLLARY 5.2. Let

$$(5.46)\qquad \overline{\theta}(h,k) = \frac{1}{k}\sum_{i=1}^{k} h(\lambda_i)X_{(n_i)}\ .$$

Then

$$(5.47) \qquad E[\overline{\theta}(h,k)] = \int_0^1 h(u)G(u)G'(u)f^*[G(u)]du$$

$$+ \theta \int_0^1 h(u)G'(u)f^*[G(u)]du + O(n^{-1})$$

and Var $[\overline{\theta}(h,k)]$ is given by (5.39).

Proof: The proof is immediate from (5.37) - (5.39) .

COROLLARY 5.3. Assume that the spacing is uniform, i.e. ,

$$(5.48) \qquad \lambda_i = \frac{i}{k+1} , \qquad i = 1,\ldots, k .$$

Then

$$(5.49) \quad E[\overline{\theta}(h,k)] = \int_0^1 h(u)G(u)du + \theta \int_0^1 h(u)du + O(n^{-1})$$

and

$$(5.50) \quad \mathrm{Var}[\overline{\theta}(h,k)]$$

$$= \frac{1}{n}\int_0^1 \int_0^1 k(u,v)h(u)h(v)G'(u)G'(v)dudv + O(n^{-1}) ,$$

as given by Jung for the full sample.

Proof: By (5.36)

$$\lambda_i = \frac{i}{k+1} = F[G^*(\frac{i}{k+1})] \ , \quad i = 1,\ldots, k$$

$$\Longrightarrow \quad G(u) \equiv G^*(u)$$

$$\Longrightarrow \quad G'(u)f^*[G(u)] \equiv 1$$

and the result follows from Corollary 5.2.

Next we have a theorem yielding the asymptotically best systematic linear estimator of θ based on a continuous weight function and arbitrary given spacing $\{\lambda_i\}$.

THEOREM 5.2. _If the assumptions of Lemma 5.2 hold and_

51) $|D| =$

$$\begin{vmatrix} \int_0^1 h_0(v)G'(v)f^*[G(v)]dv & \int_0^1 h_1(v)G'(v)f^*[G(v)]dv \\ \int_0^1 h_0(v)G(v)G'(v)f^*[G(v)]dv & \int_0^1 h_1(v)G(v)G'(v)f^*[G(v)]dv \end{vmatrix} \neq 0 \ ,$$

where

$$(5.52) \qquad h_0(v) = \frac{1}{f^*[G(v)][G'(v)]^2} \frac{d^2}{dv^2} \left[\frac{1}{G'(v)}\right]$$

$$h_1(v) = \frac{1}{f^*[G(v)][G'(v)]^2} \frac{d^2}{dv^2} \left[\frac{G(v)}{G'(v)}\right] \ ,$$

then the continuous weight function $h(u)$ that uniquely minimizes the asymptotic variance (5.39) of $\overline{\theta}(h,k)$ is given by

$$h(v) = \frac{h_o(v)}{\int_o^1 h_o(v)G'(v)f^*[G(v)]dv} . \tag{5.53}$$

Proof: We want to find h minimizing the functional

$$(5.54) \quad n \operatorname{Var}[\overline{\theta}(h,k)]$$

$$= \int_o^1 \int_o^1 k(u,v)h(u)h(v)[G'(u)]^2 [G'(v)]^2$$

$$\cdot f^*[G(u)]f^*[G(v)]dudv$$

subject to the subsidiary conditions

$$\int_o^1 h(u)G'(u)f^*[G(u)]du = 1 \tag{5.55}$$

and

$$\int_o^1 h(u)G(u)G'(u)f^*[G(u)]du = 0 .$$

The method is completely analogous to the one used in the proof of Theorem 5.1.

It is easy to check that $\overline{\theta}(h,k)$ is identical to the asymptotic form of the linear unbiased estimator obtained by Särndal (1962), p. 76. He gives the result

$$(5.56) \qquad \lim_{k \to \infty} k g_{1i} = \frac{\gamma_1'(u_i) f(u_i)}{e_{11} f^*(u_i)} ,$$

where

$$u_i = G(\lambda_i) ,$$

$$\gamma_1(x) = \frac{\partial \log f(x)}{\partial x} ,$$

$$e_{11} = E\left[\frac{\partial \log f(x)}{\partial x}\right]^2 ,$$

and g_{1i} is the coefficient of the ith sample quantile $X_{(n_i)}$, $i = 1, \ldots, k$. Substituting the continuous variable x for u_i, we obtain the correspondence of (5.56) with our weight function (5.53). We note further that, in view of Corollary 5.3, the weight function for the uniform spacing coincides with that given by Jung for the full sample.

5.3. Application to Mixture Model

We will next consider the problem of constructing an efficiency-robust analogue of $\overline{\theta}(h,k)$, i.e., an asymptotically admissible systematic linear unbiased estimator (AASLUE) of θ. The corresponding ABSLUE was obtained in Theorem 5.2 by solving for the function h(u) that minimizes Var $[\overline{\theta}(h,k)]$.

From the discussion of the method of constructing efficiency-robust estimators in Section 2.2, and in particular, from

(2.9), we see that an asymptotically admissible estimator $\overline{\theta}_G(h,k)$ may be obtained by minimizing $\mathrm{Var}[\overline{\theta}_G(h,k)|G]$, where

$$(5.57) \qquad \mathrm{Var}[\overline{\theta}_G(h,k)|G] = \sum_{\lambda=1}^{m} g_\lambda \, \mathrm{Var}[\overline{\theta}_G(h,k)|\lambda] \; .$$

The model is the usual one: we assume given m densities $f(x - \theta|\lambda)$, $\lambda = 1,\ldots, m$, a distribution $G = \{g_1,\ldots, g_m\}$, $0 < g_\lambda < 1 \;\forall\, \lambda$, and a set of k sample quantiles $X_{(n_1)},\ldots, X_{(n_k)}$. We assume that all regularity conditions introduced earlier are satisfied when necessary.

Thus we need to solve for h minimizing

$$(5.58) \quad \sum_{\lambda=1}^{m} g_\lambda \int_0^1 \int_0^1 k(u,v)h(u)h(v)[G_\lambda'(u)]^2[G_\lambda'(v)]^2 \cdot f^*[G_\lambda(u)]f^*[G_\lambda(v)]dudv \; ,$$

where

$$G_\lambda(u) = F_\lambda^{-1}(u) = x$$

and

$$G_\lambda'(u) = \frac{1}{f(x|\lambda)} \; .$$

As we see from Corollary 5.3, for the uniform spacing $\lambda_i = i/(k+1)$ the problem coincides with that of obtaining an asymptotically admissible linear unbiased estimator (AALUE) of θ, that is, the corresponding estimator for the full sample size n. Using the same weight function $\hat{h}(u)$ we have then

(5.59) $$\overline{\theta}_G(\hat{h}) = \frac{1}{n} \sum_{i=1}^{n} \hat{h}(\tfrac{i}{n+1}) X_{(i)}$$

and

(5.60) $$\overline{\theta}_G(\hat{h},k) = \frac{1}{k} \sum_{i=1}^{k} \hat{h}(\tfrac{i}{k+1}) X_{(n_i)} \ .$$

The problem of obtaining the AALUE has been solved by Yhap (1966, 1967), and is thus directly applicable to our situation for the particular spacing mentioned above.

The choice of optimum spacings for any given k will be treated in the next chapter; however, further numerical work may corroborate the early indication that a rather simple method of spacing, such as the uniform, will yield efficiencies almost equal to those of the "optimum" spacing. If an increase of precision is desired, the possibility of slightly increasing k may be considered more attractive than improvement in the spacing. Thus the estimator for the uniform spacing is of particular interest here.

In view of this fact we will briefly outline Yhap's method, but without giving explicitly here the somewhat lengthy statement of the underlying regularity conditions. The method entails minimizing with respect to h the functional

(5.61) $$\sum_{\lambda=1}^{m} g_\lambda \int_0^1 \int_0^1 k(u,v)h(u)h(v)G'(u)G'(v)dudv$$

subject to the subsidiary conditions

(5.62) $$\int_0^1 G_\lambda(u)h(u)du = \alpha_\lambda \ , \quad \lambda = 0, 1, \ldots, m \ ,$$

where

$$G_o(u) \equiv 1 \ , \quad \alpha_o = 1 \ , \quad \text{and} \quad \alpha_\lambda = 0 \quad \text{for} \quad \lambda \neq 0 \ .$$

Let

(5.63) $$\mathcal{Y} = \{J \mid J = J(u) = \sum_{\lambda=1}^{m} \gamma_\lambda G_\lambda(u) \ , \ \gamma_\lambda \text{ real}, \ \lambda = 1, \ldots, m\}$$

and define the following bilinear functional on $\mathcal{Y} \times \mathcal{Y}$:

(5.64) $$(J_i, J_j)_h = \int_0^1 \int_0^1 k(u,v) J_i'(u) J_j'(u) h(u) h(v) du dv \ ,$$

where J_i, $J_j \in \mathcal{Y}$, and $k(u,v)$ is as defined by (5.6). Yhap shows that under certain regularity conditions $(J_i, J_j)_h$ is a real inner product on $\mathcal{Y} \times \mathcal{Y}$, and then he orthonormalizes G_λ, $\lambda = 1, \ldots, m$, with respect to this inner product. The result is a set of orthonormal functions

(5.65) $$J_j(u) = \frac{G_j(u) - \sum_{i=1}^{j-1} (G_j, J_i) J_i(u)}{\|G_j(u) - \sum_{i=1}^{j-1} (G_j, J_i) J_i(u)\|} \ , \quad j = 1, \ldots, m \ ,$$

where

$$(G_j, J_i) = (G_j, J_i)_h$$

and

$$\|J\| = [(J,J)_h]^{1/2} \ .$$

Thus we can write

(5.66) $$\underline{J} = A\underline{G}$$

where

$$\underline{J} = [J_1(u), \ldots, J_m(u)]' ,$$

$$\underline{G} = [G_1(u), \ldots, G_m(u)]' ,$$

and

$$A = [\gamma_{ij}] \text{ is a non-singular matrix,}$$

so that

(5.67) $$\underline{G} = B\underline{J} ,$$

where

$$B = [b_{ij}] = A^{-1} .$$

Yhap next shows that the functional (5.61) to be minimized can be expressed in terms of these orthonormal functions; i.e.,

(5.68) $$\sum_{i=1}^{m} g_i \int_0^1 \int_0^1 G_i'(u)h(u)k(u,v)G_i'(v)h(v)dudv$$

$$= \int_0^1 \int_0^1 \sum_{p=1}^{m} c_p J_p'(u)h(u)k(u,v) \sum_{q=1}^{m} c_q J_q'(v)h(v)dudv ,$$

where

$$c_j = [\sum_{i=1}^{m} g_i b_{ij}^2]^{1/2} .$$

By a method completely analogous to that given in the proof of Theorem 5.1, we now obtain as the function minimizing (5.68),

$$(5.69) \qquad h_o(v) = \frac{1}{\sum_{q=1}^{m} c_q J_q'(v)} \frac{d^2}{dv^2} \left[\frac{1}{\sum_{p=1}^{m} c_p J_p'(v)} \right]$$

and

$$(5.70) \qquad \hat{h}(v) = \frac{h_o(v)}{\int_o^1 h_o(v)dv} .$$

The basic problem now of course is that the inner product $(J_i, J_j)_h$ with respect to which orthonormalization takes place is itself a function of h, so that we can obtain a solution only by iteration. Such an iterative procedure has been developed in detail by Yhap.

In order to generalize this approach to obtain a continuous weight function for an arbitrary $\{\lambda_i\}$, one must show that the expression suggested by (5.58), namely,

$$(5.71) \qquad \int_o^1 \int_o^1 k(u,v)h(u)h(v)[G_i'(u)]^2[G_j'(v)]^2 \cdot f^*[G_i(u)]f^*[G_j(v)]dudv ,$$

is an inner product. Orthonormalization and an iterative procedure analogous to Yhap's should then yield the desired solution. The author plans to carry out this work in the near future.

Chapter 6

THE PROBLEM OF OPTIMUM SPACING

6.1. Ogawa's Optimum Solution for a Single Shape

Fisher's Information for the asymptotic k-variate normal model defined above in (3.11) is of the form

$$K(k|L) = \sum_{i=0}^{k} \frac{[f(u_{i+1})-f(u_i)]^2}{\lambda_{i+1}-\lambda_i} , \tag{6.1}$$

where we use L to specify the given spacing, i.e.,

$$L = (\lambda_1, \ldots, \lambda_k) . \tag{6.2}$$

This expression was introduced in Section 3.4 (Eq. (3.34)) in the context of the mixture model as $K_\lambda(k)$, the Information corresponding to shape λ. There we also saw the relationship

$$K(k|L) = 1'S1 = \frac{1}{n \operatorname{Var}[\tilde{\theta}(k)]} , \tag{6.3}$$

where

$$\tilde{\theta}(k) = \frac{1'S\underline{x}^{(k)}{}'}{1'S1}$$

is the ABSLUE of θ (as shown in Section 4.3), the estimator studied by Ogawa (1951).

Ogawa defined as optimum spacing the vector $L = (\lambda_1, \ldots, \lambda_k)$ that maximizes (6.1), the Information with respect to θ

(or, equivalently, minimizes the variance of the ABSLUE of θ).

Differentiating (6.1) with respect to λ_i, we obtain

$$(6.4)\quad \frac{\partial K(k|L)}{\partial \lambda_i} = \left(\frac{f_{i+1}-f_i}{\lambda_{i+1}-\lambda_i} - \frac{f_i-f_{i-1}}{\lambda_i-\lambda_{i-1}}\right)\left(\frac{f_{i+1}-f_i}{\lambda_{i+1}-\lambda_i} + \frac{f_i-f_{i-1}}{\lambda_i-\lambda_{i-1}} - \frac{2f_i'}{f_i}\right) = 0 ,$$

$$i = 1,\ldots, k ,$$

where

$$f_i = f(u_i) , \quad f_i' = \frac{df(u)}{du}\Big|_{u=u_i} .$$

Tischendorf (1955) gave regularity conditions for the pdf f under which a necessary condition for L to maximize (6.1) is that

$$(6.5)\qquad \frac{f_{i+1}-f_i}{\lambda_{i+1}-\lambda_i} + \frac{f_i-f_{i-1}}{\lambda_i-\lambda_{i-1}} - \frac{2f_i'}{f_i} = 0 , \quad i = 1,\ldots, k .$$

This had already been shown for the normal distribution by Ogawa, who also proved that in this case the system of equations (6.5) has one and only one solution and that the optimum spacing is symmetric (cf. Ogawa in Sarhan and Greenberg (1962), p. 273).

Numerical solutions for the normal case were obtained by Ogawa for $k \leq 10$, and the resulting relative efficiencies $E(k|L)$, where

(6.6) $$E(k|L) = \frac{K(k|L)}{I}, \quad I = E[\frac{\partial \log f(x)}{\partial x}]^2,$$

were compared with those obtained by using the approximate spacing

(6.7) $$\lambda_i = \frac{i - \frac{1}{2}}{k} .$$

Ogawa also computed the best spacing for the exponential distribution for $k \leq 15$. All these results are tabulated in Sarhan and Greenberg (1962).

Särndal (1962) gives $\lambda_i = i/(k+1)$ as the optimum spacing for the logistic distribution. A proof of this and related numerical results are found in Gupta and Gnanadesika (1966).

Optimum spacings do not seem to be known for other families of distributions. It is interesting to observe from these results that (a) the approximate spacing (6.7) compares very favorably with the optimum one, and (b) high efficiencies ($\geq .9$) are obtained already for small values of $k (4 \leq k \leq 10)$.

6.2. Särndal's Nearly Optimum Solution for a Single Shape

In view of the extreme difficulty of solving for the optimum spacing, Särndal (1962) proposed an approximate solution expressed in terms of the generating function $f^*(x)$

defined in Section 5.2. His nearly optimum spacing, or equivalently, the generating function, is obtained by using an asymptotic expansion and a limiting integral involving the relative efficiency $E(k|L)$ defined in (6.6) and by calculus of variations solving for the function $f^*(x) = f_m^*(x)$ that maximizes the efficiency.

Briefly, following Särndal (Chapter 4), we have

$$(6.8) \qquad E(k|L) = \frac{K(k|L)}{I} = 1 - \frac{I-K(k|L)}{I} .$$

Expanding both I and $K(k|L)$ in Taylor series, Särndal obtains for the difference

$$(6.9) \quad I - K(k|L)$$

$$= \frac{1}{12} \sum_{i=0}^{k} (\lambda_{i+1}-\lambda_i)^3 \{f''[G(\lambda_i)]\}^2 + O[(\lambda_{i+1}-\lambda_i)^4] .$$

Further, we have the relationship

$$(6.10) \qquad \lambda_{i+1} - \lambda_i = F[G^*(\pi_{i+1})] - F[G^*(\pi_i)]$$

$$= \frac{1}{k+1} \frac{f(u_i)}{f^*(u_i)} + O(k^{-2}) .$$

This follows from the (extended) mean value theorem for $a = \pi_i$, $b = \pi_{i+1}$, since

$$(6.11) \qquad \left.\frac{dF}{dt}\right|_{t=\pi_i} = \frac{f(u_i)}{f^*(u_i)} .$$

Substituting (6.9) and (6.10) in (6.8), we obtain

$$(6.12)\quad E(k|L) = 1 - \frac{1}{12I(k+1)^3} \sum_{i=0}^{k} \left[\frac{f(u_i)}{f^*(u_i)}\right]^3 \{f''[G(\lambda_i)]\}^2 + O(k^{-4})$$

$$= 1 - \frac{1}{12I(k+1)^3} \sum_{i=0}^{k} \frac{[\gamma'(u_i)]^2 f(u_i)}{[f^*(u_i)]^3} + O(k^{-4}) \ ,$$

(where $\gamma(x) = \frac{\partial \log f(x)}{\partial x}$)

$$= 1 - \frac{1}{12I(k+1)^2} \int_{-\infty}^{\infty} \frac{[\gamma'(x)]^2 f(x)}{[f^*(x)]^2}\, dx + o(k^{-2}),$$

if the integral exists.

Up to this point we have assumed that all necessary regularity conditions on f are satisfied, and the integral is obtained by arguments similar to those given in the proof of Lemma 5.1. If the integral in (6.12) does not exist, Särndal considers the alternative expression

$$(6.13)\quad E(k|L,\varepsilon) = 1 - \varepsilon - \frac{1}{12I(k+1)^2}\left[\int_{G(\delta)}^{G(1-\delta)} \frac{[\gamma'(x)]^2 f(x)}{[f^*(x)]^2} + \varepsilon_1\right]$$

$$+ o(k^{-2}) \ ,$$

with $\varepsilon > 0$, ε_1 arbitrarily small and $\delta = \delta(\varepsilon)$ small and positive.

Thus by (6.12) or (6.13) the relative efficiency is expressed explicitly as a function of k and the spacing λ_i

(via $f^*(x)$). We see that:

(a) For fixed $f^*(x)$, we can increase the efficiency by increasing k; the efficiency tends to 1 (6.12) or $1 - \varepsilon$ (6.13). ε is the loss of efficiency caused by cutting off the extremes of the range.

(b) For fixed k, we can maximize the efficiency by a choice of $f^*(x)$ that minimizes the integral in (6.12) or (6.13).

In either case the integral is minimized subject to the subsidiary condition $\int_{-\infty}^{\infty} f^*(x)dx = 1$.

Straightforward application of the calculus of variations yields

$$f_m^*(x) = \frac{1}{c}\{[\gamma'(x)]^2 f(x)\}^{1/3} , \tag{6.14}$$

where

$$c = \int_{-\infty}^{\infty} \{[\gamma'(x)]^2 f(x)\}^{1/3} dx .$$

Särndal classifies $f_m^*(x)$ as Type I if the integral in (6.12) exists; otherwise it is said to be of Type II. It evidently yields a nearly optimum solution since it is based on an approximate expression for the efficiency $E(k|\lambda)$, but the latter still tends to at least $1 - \varepsilon$ as $k \longrightarrow \infty$, where ε is arbitrarily small.

The nearly optimum generating function was obtained by Särndal for a number of distributions, including the cases studied numerically by Ogawa. Tabulated comparisons of relative efficiencies for different values of k indicate good results for these nearly optimum spacings.

In a number of instances $f_m(x)$ takes a simple form. For example, for the normal distribution with only the mean θ unknown, we have

$$f_m^*(x) = \frac{1}{\sqrt{6\pi}} e^{-x^2/6} ,$$

which generates the nearly optimum spacing

$$\begin{aligned} \lambda_i &= F[G_m^*(\pi_i)] \\ &= \Phi[\sqrt{3}\, \Phi^{-1}(\pi_i)] , \qquad i = 1, \ldots, k . \end{aligned}$$

Särndal remarks that such simple expressions for the generating function are the exception and not the rule, but points out that even numerical integration and tabulation are in general easier to perform than solving the system of equations (6.4) for the optimum spacing, which, moreover, has to be done separately for each value of k.

6.3. Admissibility Concept for Mixture Model

Given a family of m shapes $f(x|\lambda)$, λ, $j \in \Lambda$, define, as in (3.34)

(6.15) $$K_\lambda(k|L) = \sum_{i=0}^{k} \frac{[f_{\lambda,i+1} - f_{\lambda,i}]^2}{\lambda_{i+1} - \lambda_i} ,$$

where

$$f_{\lambda,i} = f(u_i^\lambda|\lambda)$$

and

$$u_i^\lambda = G_\lambda(\lambda_i) , \qquad i = 1,\ldots, k .$$

Furthermore, let

(6.16) $$E_\lambda(k|L) = \frac{K_\lambda(k|L)}{I(\lambda)} ,$$

where

$$I(\lambda) = E[\frac{\partial \log f(x|\lambda)}{\partial x}]^2 .$$

For any k the optimum spacing for shape j, say

(6.17) $$L_j^o = (\lambda_{1j}^o,\ldots, \lambda_{kj}^o) ,$$

is obtained by maximizing (6.15), yielding

(6.18) $$E_j(k|L^o) = \frac{K_j(k|L_j^o)}{I(j)} ,$$

the efficiency of the optimum spacing L_j^o. Then the relative efficiency of any other spacing is given by

(6.19) $$Re_j(k|L,L^o) = \frac{E_j(k|L)}{E_j^o(k|L^o)} = \frac{K_j(k|L)}{K_j(k|L_j^o)} .$$

We are interested in obtaining a spacing L which in some sense is optimum, say admissible, over the m shapes $f(x|j)$, $j \in \Lambda$.

LEMMA 6.1. A spacing $L^* = (\lambda_1^*, \ldots, \lambda_k^*)$ is admissible over Λ if it maximizes $\sum_{j=1}^{m} g_j K_j(k|L)$, i.e., if

$$(6.20) \qquad K_G(k|L^*) = \sum_{j=1}^{m} g_j K_j(k|L^*) = \sup_L \sum_{j=1}^{m} g_j K_j(k|L) .$$

Proof: Assume L^* is not admissible over j. Then $\exists$ a spacing $L' \ni$

$$K_j(k|L') \geq K_j(k|L^*) \quad \forall j$$

with inequality strict for at least one j. But then

$$\sum_{j=1}^{m} g_j K_j(k|L') > \sum_{j=1}^{m} g_j K_j(k|L^*) ,$$

contradicting (6.20).

This definition of a spacing admissible over Λ applies analogously to the nearly optimum solution proposed by Särndal.

Regardless of how L^* was obtained, it can be evaluated by the expression

$$(6.21) \qquad Re_j(k|L^*, L_j^o) = \frac{K_j(k|L^*)}{K_j(k|L_j^o)} \leq 1 .$$

As pointed out in Section 6.1, for the case of a single shape, maximizing $K(k|L)$ is equivalent to minimizing the variance of the corresponding estimator (ABSLUE) of θ. This analogy does not hold for the mixture model, since in general

(6.22) $$\sum_{j=1}^{m} g_j K_j(k|L) \neq 1'[\sum_{j=1}^{m} g_j \Gamma_j]^{-1} 1 ,$$

where the second term is the reciprocal of the variance (times n) of the corresponding AASLUE, as defined in Theorem 4.5.

Here we deal exclusively with the concept of Information based on the k-variate normal model, and consider it in various forms. This approach is applicable to other estimators, such as the systematic Pitman estimator discussed in Chapter 3, for which also a spacing must be provided.

In practice usual computational procedures can be applied to determine admissible and maximin estimators, with the class of estimators extended so as to include those based on all possible spacings.

6.4. Admissible Spacing Based on Exact Solution

The problem is to solve for L maximizing

(6.23) $$K_G(k|L) = \sum_{j=1}^{m} g_j K_j(k|L) .$$

A necessary condition is that L satisfy the system of equations

$$(6.24)\quad \frac{\partial K_G(k|L)}{\partial \lambda_i} = \sum_{j=1}^{m} g_j \frac{\partial K_j(k|L)}{\partial \lambda_i}$$

$$= \sum_{j=1}^{m} g_j \left[\left(\frac{f_{i+1,j}-f_{ij}}{\lambda_{i+1}-\lambda_i} - \frac{f_{ij}-f_{i-1,j}}{\lambda_i-\lambda_{i-1}}\right) \cdot \left(\frac{f_{i+1,j}-f_{ij}}{\lambda_{i+1}-\lambda_i} + \frac{f_{ij}-f_{i-1,j}}{\lambda_i-\lambda_{i-1}} - \frac{2f'_{ij}}{f_{ij}}\right)\right] = 0,$$

$$i = 1,\ldots, k .$$

The problem of how to solve for L (questions concerning the existence of any simplifying necessary conditions, uniqueness of the solution, or ways of finding the maximizing set) remains open.

Evidently this is a much more complex situation than the one involving only a single shape. Writing (6.4) simply as

$$(6.25)\qquad x_i(L)y_i(L) = 0 , \quad i = 1,\ldots, k ,$$

Tischendorf's regularity conditions on f guarantee that $x_i(L) \neq 0$ for any i, L, and this yields the necessary condition (6.5), i.e.,

$$(6.26)\qquad y_i(L) = 0 , \quad i = 1,\ldots, k.$$

The corresponding system of equations for the mixture model is

$$(6.27) \qquad \sum_{j=1}^{m} g_j x_{ij}(L) y_{ij}(L) = 0 ,$$

and assuming that $x_{ij}(L) \neq 0$ for any i, j, L does not yield any immediate simplification, since $x_{ij}(L)$ and $y_{ij}(L)$ may both assume positive and negative values.

Examining this system for specific families of shapes and small values of k would be a start in attacking the problem, unless a large-scale numerical investigation were desired.

6.5. Another Possible Approach

Given the optimum spacing L_j^o for each shape $j \in \Lambda$, it is possible and plausible that the following relationship holds:

$$(6.28) \qquad \min_j \lambda_{ij}^o \leq \lambda_i^* \leq \max_j \lambda_{ij}^o , \qquad i = 1, \ldots, k ,$$

where

$$(6.29) \qquad K_j(k|L_j^o) = \sup_L K_j(k|L)$$

and

$$(6.30) \qquad K_G(k|L^*) = \sup_L \sum_{j=1}^{m} g_j K_j(k|L) ;$$

$L^* = (\lambda_1^*, \ldots, \lambda_k^*)$ is the admissible spacing corresponding to G.

If (6.28) is true and L_j^o is actually known $\forall$ j, a suitable approximation to L* would be given by

(6.31) $$L_G = \sum_{j=1}^{m} g_j L_j^o ,$$

i.e.,

$$\lambda_{i,G} = \sum_{j=1}^{m} g_j \lambda_{ij}^o , \quad i = 1,\ldots, k ,$$

which clearly satisfies (6.28):

(6.32) $$\min_j \lambda_{ij}^o \leq \lambda_{i,G} \leq \max_j \lambda_{ij}^o , \quad i = 1,\ldots, k .$$

Briefly the problem might be stated as follows:

Given a set of real-valued functions $K_j(L)$ whose argument is a vector $L = (\lambda_1,\ldots, \lambda_k)$, such that $0 < \lambda_1 < \ldots < \lambda_k < 1$, let L_j be the vector maximizing the function $K_j(L)$. Under what conditions is it true that the vector L* maximizing a given convex combination of the $K_j(L)$ is itself contained (component-wise) in the convex hull of the L_j^o?

An example. Consider the example of two symmetric shapes, the normal and logistic distributions, for k = 3. We have $\lambda_3 = 1 - \lambda_1$, $\lambda_2 = .5$, so that we need find λ_1 only. We know that

$$\lambda_{11}^o = .163 \quad \text{(normal)}$$

$$\lambda_{12}^o = .25 \quad \text{(logistic)}$$

and here we can show numerically that

$$.163 \le \lambda_1^*(g) \le .25$$

by considering solutions of (cf. Eq. (6.24))

$$g\left(\frac{.4-f_1}{.5-\lambda_1} - \frac{f_1}{\lambda_1}\right)\left(\frac{.4-f_1}{.5-\lambda_1} + \frac{f_1}{\lambda_1} + 2u_1\right) + (1-g)(.25-\lambda_1) = 0 \ .$$

As g varies from 1 to 0, λ_1^* sweeps out all values from .163 to .25.

6.6. Admissible Spacing Based on Approximate Solution

Applying the concept of an admissible spacing to Särndal's nearly optimum solution, we might consider maximizing

$$(6.33) \quad E_G(k|L) = \sum_{j=1}^{m} g_j E_j(k|L) = \sum_{j=1}^{m} g_j \frac{K_j(k|L)}{I(j)}$$

$$= \sum_{j=1}^{m} g_j \left[1 - \frac{I(j)-K_j(k|L)}{I(j)}\right]$$

$$= 1 - \sum_{j=1}^{m} \frac{g_j}{I(j)} [I(j)-K_j(k|L)]$$

$$= 1 - \sum_{j=1}^{m} \frac{g_j}{I(j)} \left[\frac{1}{12} \sum_{i=0}^{k} (\lambda_{i+1}-\lambda_i)^3 \{f_j''[G_j(\lambda_i)]\}^2\right.$$

$$\left. + O[(\lambda_{i+1}-\lambda_i)^4]\right] ,$$

where we have made use of (6.9), adding the subscript for shape j.

Here (6.10) is of the form

$$(6.34) \qquad \lambda_{i+1}-\lambda_i = \frac{1}{k+1}\,\frac{f_j(u_i^j)}{f_j^*(u_i^j)} + O(k^{-2}) \; ,$$

where

$$u_i^j = G_j(\lambda_i) \; ,$$

so that (6.33) becomes

$$(6.35) \quad E_G(k|L) = 1 - \frac{1}{12(k+1)^3}\sum_{j=1}^{m}\frac{g_j}{I(j)}\sum_{i=0}^{k}\left[\frac{f_j(u_i^j)}{f_j^*(u_i^j)}\right]^3 [f_j''(u_i^j)]^2 + O(k^{-4})$$

$$= 1 - \frac{1}{12(k+1)^3}\sum_{j=1}^{m}\frac{g_j}{I(j)}\sum_{i=0}^{k}\frac{[\gamma_j'(u_i^j)]^2 f_j(u_i^j)}{[f_j^*(u_i^j)]^3} + O(k^{-4})$$

(where $\gamma_j(x) = \dfrac{\partial \log f_j(x)}{\partial x}$)

$$= 1 - \frac{1}{12(k+1)^3}\sum_{j=1}^{m}\frac{g_j}{I(j)}\int_{-\infty}^{\infty}\frac{[\gamma_j'(x)]^2 f_j(x)}{[f_j^*(x)]^2}\,dx + o(k^{-2}) \; ,$$

if the integral exists $\forall j$.

If the integral does not exist, an expression analogous to (6.13) may be defined.

Next we need to minimize the integral

$$\int_{-\infty}^{\infty} \sum_{j=1}^{m} \frac{g_j[\gamma_j'(x)]^2 f_j(x)}{I(j)[f_j^*(x)]^2}\, dx \ , \tag{6.36}$$

but here the correspondence with the single-shape problem breaks down, since we are dealing with m generating functions $f_j^*(x)$.

Assuming that a set of functions $f_j^*(x)$ were found that jointly minimizes (6.36), this set would yield, by definition,

$$\lambda_{ij} = F_j[G_j^*(\pi_i)] \ , \quad i = 1,\ldots, k \ , \tag{6.37}$$

i.e., m distinct spacings L_j, whereas we need one spacing to determine the sample quantiles to be used in the estimator.

One way to get around this problem is to allow only one generating function to be independently selected which then determines the others.

Without loss of generality, assume that $f_1^*(x)$ determines the spacing. Then

$$\lambda_i = F_1[G_1^*(\pi_i)] \ , \quad i = 1,\ldots, k \ ; \tag{6.38}$$

but let also

$$\lambda_i = F_j[G_j^*(\pi_i)] \ , \quad j = 2,\ldots, m \ , \tag{6.39}$$

so that

$$G^*_j(\pi_i) = G_j\{F_1[G^*_1(\pi_i)]\} \ , \quad i = 1,\ldots, k, \ j = 1,\ldots, m,$$

from which we may obtain $f^*_j(x) = \frac{1}{G^{*\prime}(u)}$. Substituting

$$(6.40) \qquad f^*_j(x) = \frac{f_j\{f_1[\frac{1}{f^*_1(x)}]\}f^*_1(x)}{f_1[\frac{1}{f^*_1(x)}]}$$

in (6.36), we obtain the result that the "nearly" admissible spacing L* is generated by the function $f^*_1(x)$ that minimizes the functional

$$(6.41) \qquad \int_{-\infty}^{\infty} \sum_{j=1}^{m} \frac{g_j[\gamma'_j(x)]^2\{f_1[\frac{1}{f^*_1(x)}]\}^2 dx}{I(j)f_j[f_1(\frac{1}{f^*_1(x)})]\cdot[f^*_1(x)]^2}$$

subject to the subsidiary condition

$$(6.42) \qquad \int_{-\infty}^{\infty} f^*_1(x)dx = 1 \ .$$

The Euler equation corresponding to this problem cannot be solved explicitly for $f^*_1(x)$, so that an iterative procedure needs to be developed.

6.7. Discussion

As is evident in Sections 6.4-6.6, complete solution of the problem of admissible spacing requires additional investigation even in the realm of theory. The author intends to pursue some of these questions further in the future.

In addition, large-scale numerical investigations are in order. As was pointed out in Section 5.3, new numerical results may continue to provide evidence in favor of settling on simple spacings, and thus may obviate the necessity of searching for optimum solutions. (Plausible criteria for evaluating a spacing $\{\lambda_i\}$ for each shape $j \in \Lambda$ have been defined in (6.16) and (6.19) above.) Some relevant results for small sample sizes are given in Birnbaum and Laska (1967b).

REFERENCES

1. Bennett, C. A. (1952). Asymptotic properties of ideal linear estimators. Doctoral dissertation, University of Michigan.
2. Bernstein, S. N. (1917). Theory of Probability (in Russian).
3. Birnbaum, A. (1961). Some theory and techniques for robust estimation. (Abstract) Ann. Math. Statist. 32: 622.
4. Birnbaum, A. and E. Laska (1967a). Optimal robustness: a general method, with applications to linear estimators of location. Submitted to J. Amer. Statist. Assoc.
5. Birnbaum, A. and E. Laska (1967b). Optimally robust linear estimators of location. Technical Report, Courant Institute of Mathematical Sciences, New York University. Abstract to appear in Ann. Math. Statist. (1967).
6. Blom, G. (1958). Statistical Estimates and Transformed Beta-Variables. Wiley, New York.
7. Chernoff, H. (1956). Large sample theory: parametric case. Ann. Math. Statist. 27: 1-22.
8. Chernoff, H., J. L. Gastwirth, and M. V. Johns, Jr. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. Ann. Math. Statist. 38: 52-72.
9. Craig, L. (1967). A property of bisymmetric matrices. To be submitted for publication.

10. Daniell, L. (1920). Observations weighted according to order. Amer. J. Math. 42: 222-236.

11. Eisenberger, I. and E. C. Posner (1965). Systematic statistics used for data compression in space telemetry. J. Amer. Statist. Assoc. 60: 97-133.

12. Elfving, G. (1966). Robustness of Bayes decisions against the choice of prior. Technical Report No. 122, Department of Statistics, Stanford University.

13. Girshick, M. and L. J. Savage (1951). Bayes and minimax estimates for quadratic loss functions. Proc. Second Berkeley Symp. Math. Statist. Prob. pp. 53-73, University of California Press.

14. Gupta, S. S. and M. Gnanadesikan (1966). Estimation of the parameters of the logistic distribution. Biometrika 53: 565-570.

15. Hájek, J. (1962). Asymptotically most powerful rank order tests. Ann. Math. Statist. 33: 1124-1147.

16. Hartigan, J. A. (1965). The asymptotically unbiased prior distribution. Ann. Math. Statist. 36: 1137-1152.

17. Huber, P. (1964). Robust estimation of a location parameter. Ann. Math. Statist. 35: 73-101.

18. Jung, J. (1955). On linear estimates defined by a continuous weight function. Arkiv för Matematik, Band 3, Nr. 15, 199-209.

19. Karlin, S. (1958). Admissibility for estimation with quadratic loss. Ann. Math. Statist. 29: 406-436.

20. Laplace, P. S. de (1820). Théorie Analytique des Probabilités (3rd edition). Paris.

21. Laska, E. (1962). A general theory of robustness. Doctoral dissertation, New York University.

22. Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. University of California Publications in Statistics 1: 277-330.

23. Le Cam, L. (1958). Les propriétés asymptotiques des solutions de Bayes. Publ. Inst. Statist. Univ. Paris. 2: 17-35.

24. Lindley, D. V. (1961). The use of prior probability distributions in statistical inference and decisions. Proc. Third Berkeley Symp. Math. Statist. Prob. 1: 453-468.

25. Lindley, D. V. (1965). Introduction to Probability and Statistics from a Bayesian Viewpoint. Cambridge University Press.

26. Lloyd, E. H. (1952). Least-squares estimation of location and scale parameters using order statistics. Biometrika 39: 88-95.

27. Mises, R. von (1919). Fundamentalsätze der Wahrscheinlichkeitsrechnung. Math. Zeit. 4: 1-97.

28. Mosteller, F. (1946). On some useful "inefficient" statistics. Ann. Math. Statist. 17: 377-408.

29. Ogawa, J. (1951). Contributions to the theory of systematic statistics, I. Osaka Math. J. 3: 175-213.

30. Pitman, E. J. G. (1939). The estimation of location and scale parameters of a continuous population of any form. Biometrika 30: 391-421.

31. Rao, C. R. (1965). Linear Statistical Inference and Its Applications. Wiley, New York.

32. Sarhan, A. E. and B. G. Greenberg (1962). Contributions to Order Statistics. Wiley, New York.

33. Särndal, C. E. (1962). Information from Censored Samples. Almqvist and Wiksell, Stockholm.

34. Stein, C. (1956). Efficient nonparametric testing and estimation. Proc. Third Berkeley Symp. Math. Statist. Prob. 1: 187-195.

35. Stein, C. (1959). The admissibility of Pitman's estimator of a single location parameter. Ann. Math. Statist. 30: 970-979.

36. Tischendorf, J. A. (1955). Linear estimation techniques using order statistics. Doctoral dissertation, Purdue University.

37. Tukey, J. (1960). A survey of sampling from contaminated distributions. Contributions to Probability and Statistics (I. Olkin et al, eds.) pp. 448-485, Stanford University Press.

38. Tukey, J. (1962). The future of data analysis. Ann. Math. Statist. 33: 1-67.

39. Wilks, S. S. (1963). Mathematical Statistics (2nd printing). Wiley, New York.

40. Wolfowitz, J. (1953). The method of maximum likelihood and the Wald theory of decision functions. Indag. Math. 15: 114-119.

41. Yhap, E. F. (1966). Asymptotically admissible linear unbiased estimators (AALUE) for robust estimation of location and scale parameters. (Abstract) Ann. Math. Statist. 37: 1071-1072.

42. Yhap, E. F. (1967). An asymptotically optimally robust linear unbiased estimator of location for symmetric shapes. Doctoral dissertation, New York University.

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Courant Institute of Mathematical Sciences New York University	2a. REPORT SECURITY CLASSIFICATION Not classified 2b. GROUP None

3. REPORT TITLE

Contributions to Robust Estimation

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Technical Report September 1967

5. AUTHOR(S) *(Last name, first name, initial)*

Miké, Valerie

6. REPORT DATE September	7a. TOTAL NO. OF PAGES 125	7b. NO. OF REFS 42
8a. CONTRACT OR GRANT NO. Nonr-285(38) b. PROJECT NO. NR 042-206 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) IMM 361 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* None	

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY U.S. Navy, Office of Naval Research 207 West 24th St., New York, N.Y.

13. ABSTRACT

For the problem of efficiency-robust estimation of the location parameter θ of a family of symmetric pdf's $f(x-\theta|\lambda)$, $\lambda \epsilon \Lambda = \{1,\ldots,m\}$, $\theta \epsilon \Theta = \{\theta | -\infty < \theta < \infty\}$, the method of "mixture models" of Birnbaum (Ann. Math. Statist. 32 (1961) 622) is applied to determine generalized Pitman estimators, which are shown to be admissible, with squared error loss function, under broad regularity conditions. With increasing sample size, these estimators are proved to be fully efficient (i.e., asymptotically equivalent, for each value of λ, to the maximum likelihood estimator which would be appropriate if the true value of λ were known). Computationally tractable analogous estimators based on k sample quantiles are defined in the context of the model representing their asymptotic normal distributions. It is shown that with increasing k these approach equivalence to the fully efficient estimators based on complete samples. Equivalent estimators are given also for the case of unknown scale parameters. Efficiency-robust linear unbiased estimators based on sample quantiles are derived and the optimal spacing of quantiles is discussed.

DD FORM 1 JAN 64 1473

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, *or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

BASIC DISTRIBUTION LIST FOR UNCLASSIFIED TECHNICAL REPORTS

Head, Logistics and Mathematical Statistics Branch Office of Naval Research Washington, D.C. 20360	3 copies
Commanding Officer Office of Naval Research Branch Office Navy #100 Fleet Post Office New York, New York	2 copies
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20 copies
Defense Logistics Studies Information Exchange Army Logistics Management Center Fort Lee, Virginia Attn: William B. Whichard	1 copy
Technical Information Officer Naval Research Laboratory Washington, D.C. 20390	6 copies
Commanding Officer Office of Naval Research Branch Office 207 West 24th Street New York, New York 10011 Attn: J. Laderman	1 copy
Commanding Officer Office of Naval Research Branch Office 1030 East Green Street Pasadena 1, California Attn: Dr. A.R. Laufer	1 copy
Bureau of Supplies and Accounts Code OW Department of the Navy Washington 25, D.C.	1 copy
Institute for Defense Analyses Communications Research Division von Neumann Hall Princeton, New Jersey	1 copy
University of Chicago Statistical Research Center Chicago, Illinois Attn: Prof. Paul Meier	1 copy

Stanford University
Applied Mathematics & Statistics Lab.
Stanford, California
Attn: Prof. Gerald J. Lieberman

Florida State University
Department of Statistics
Tallahassee, Florida
Attn: Prof. I.R. Savage

Florida State University
Department of Statistics
Tallahassee, Florida
Attn: Dr. Ralph A. Bradley

Princeton University
Department of Mathematics
Princeton, New Jersey

Columbia University
Department of Mathematical Statistics
New York 27, New York
Attn: Prof. T.W. Anderson

University of California
Department of Statistics
Berkeley 4, California
Attn: Prof. J. Neyman

University of Washington
Department of Mathematics
Seattle 5, Washington
Attn: Prof. Z.W. Birnbaum

Cornell University
Department of Mathematics
Ithaca, New York
Attn: Prof. J. Wolfowitz

Harvard University
Department of Statistics
Cambridge, Massachusetts
Attn: Prof. W.G. Cochran

The Research Triangle Institute
Statistics Research Division
505 West Chapel Hill Street
Durham, North Carolina
Attn: Dr. Malcolm R. Leadbetter

Columbia University
Dept. of Industrial Engineering
New York 27, New York
Attn: Prof. Cyrus Derman

Columbia University
Department of Mathematics
New York 27, New York
Attn: Prof. H. Robbins

New York University
Courant Inst. of Mathematical Sciences
New York , New York 10012
Attn: Prof. W.M. Hirsch

Cornell University
Department of Plant Breeding
Biometrics Unit
Ithaca, New York
Attn: Walter T. Federer

University of North Carolina
Statistics Department
Chapel Hill, North Carolina
Attn: Prof. Walter L. Smith

Stanford University
Department of Statistics
Stanford, California
Attn: Prof. Herman Rubin

Brown University
Division of Applied Mathematics
Providence 12, Rhode Island
Attn: Prof. M. Rosenblatt

New York University
Department of Industrial Engineering
and Operations Research
Bronx 63, New York
Attn: Prof. J. H. Kao

University of Wisconsin
Department of Statistics
Madison, Wisconsin
Attn: Prof. G.E.P. Box

Prof. Anders Hald
Institute of Mathematical Statistics
University of Copenhagen
Copenhagen, Denmark

Yale University
Department of Mathematics
New Haven, Connecticut
Attn: Prof. L.J. Savage

Massachusetts Institute of Technology
Department of Electrical Engineering
Cambridge, Massachusetts
Attn: Dr. R.A. Howard

Stanford University
Department of Statistics
Stanford, California
Attn: Prof. E. Parzen

Arcon Corporation
803 Massachusetts Avenue
Lexington 73, Massachusetts
Attn: Dr. Arthur Albert

University of California
Institute of Engineering Research
Berkeley 4, California
Attn: Prof. R.E. Barlow

Michigan State University
Department of Statistics
East Lansing, Michigan
Attn: Prof. J. Gani

Dr. Jack R. Borsting
Chairman
Dept. of Operations Analysis
U. S. Naval Postgraduate Sch.
Monterey, California 93940

Prof. Harold E. Dodge
Rutgers - The State University
Statistics Center
New Brunswick, New Jersey

Yale University
Department of Statistics
New Haven, Connecticut
Attn: Prof. F.J. Anscombe

Purdue University
Division of Mathematical Sciences
Lafayette, Indiana
Attn: Prof. S.S. Gupta

Cornell University
Department of Industrial Engineering
Ithaca, New York
Attn: Prof. Robert Bechhofer

Stanford University
Department of Statistics
Stanford, California
Attn: Prof. C. Stein

Applied Mathematics and Statistics Lab.
Department of Statistics
Stanford University
Stanford, California
Attn: Prof. H. Solomon

Decision Studies Group
460 California Avenue
Palo Alto, California
Attn: Warren R. Ketler

The Johns Hopkins University
Department of Mathematical Statistics
34th and Charles Streets
Baltimore 18, Maryland
Att: Prof. Geoffrey S. Watson

NYU c.1
IMM-
361 Miké

Contributions to robust estimation.

NYU c.1
IMM-
361 Miké
AUTHOR
Contributions to robust
TITLE
estimation.

DATE DUE	BORROWER'S NAME
[illegible]	~~J. Z[illegible]~~
~~NOV 8 19[illegible]~~	~~Martin Feigenbaum~~
~~JAN 3 1978~~	~~Coppage~~

N.Y.U. Courant Institute of Mathematical Sciences
251 Mercer St.
New York, N. Y. 10012

NOV 17 1967

DATE DUE

[illegible]			
NOV 8 1971			
~~JAN 3 1979~~			

GAYLORD PRINTED IN U.S.A.